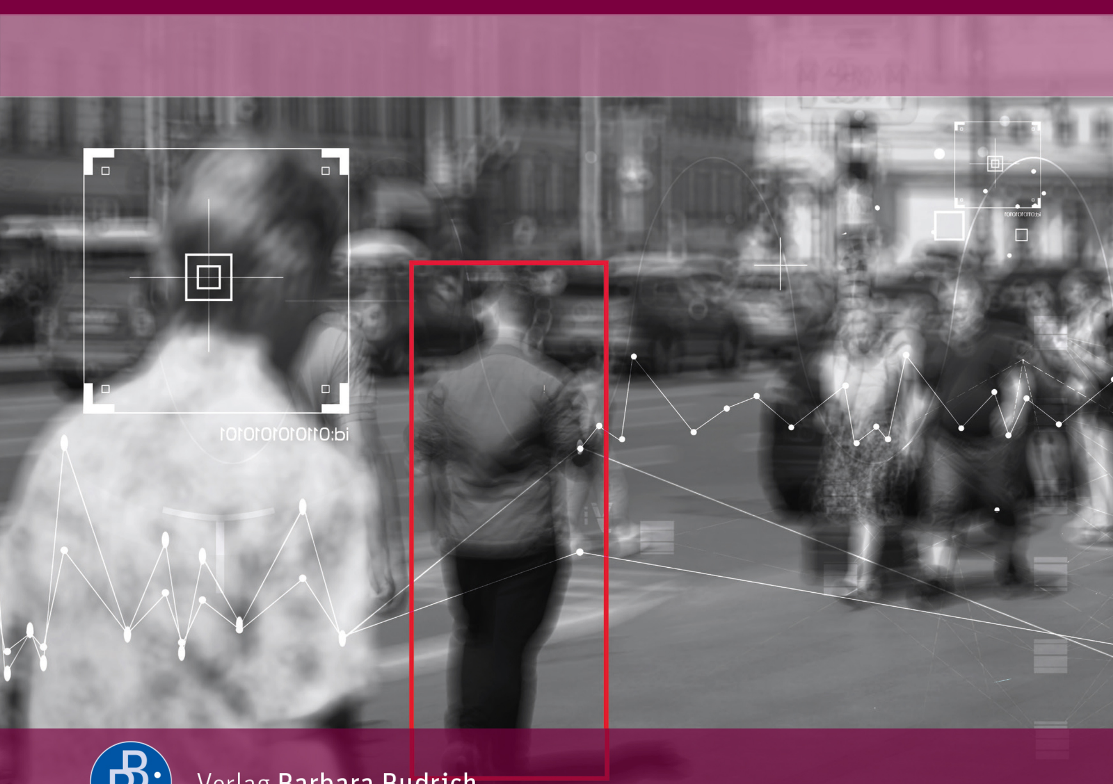
Dan Verständig, Christina Kast,
Janne Stricker, Andreas Nürnberger (Hrsg.)

Algorithmen und Autonomie

Interdisziplinäre Perspektiven auf das Verhältnis von Selbstbestimmung und Datenpraktiken



Verlag Barbara Budrich

Algorithmen und Autonomie

Dan Verständig
Christina Kast
Janne Stricker
Andreas Nürnberger (Hrsg.)

Algorithmen und Autonomie

Interdisziplinäre Perspektiven auf das Verhältnis von Selbstbestimmung und Datenpraktiken

Verlag Barbara Budrich
Opladen • Berlin • Toronto 2022

Bibliografische Information der Deutschen Nationalbibliothek. Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über https://portal.dnb.de abrufbar.

www.budrich.de



Dieses Buch steht im Open-Access-Bereich der Verlagsseite zum kostenlosen Download bereit (https://doi.org/10.3224/84742520).

Eine kostenpflichtige Druckversion (Print on Demand) kann über den Verlag bezogen werden. Die Seitenzahlen in der Druck- und Onlineversion sind identisch.

ISBN 978-3-8474-2520-5 (Paperback)
eISBN 978-3-8474-1720-0 (PDF)
DOI 10.3224/84742520

Umschlaggestaltung: Bettina Lehfeldt, Kleinmachnow – www.lehfeldtgraphic.de
Titelbildnachweis: shutterstock.com
Satz: 3w+p GmbH, Rimpar
Druck: docupoint GmbH, Barleben

Inhalt

Algorithmen und Autonomie

Interdisziplinäre Perspektiven auf das Verhältnis von Selbstbestimmung und Datenpraktiken

Dan Verständig, Christina Kast, Janne Stricker, Andreas Nürnberger

Einleitung

„An algorithm must be seen to be believed."
(Knuth 1968)

Es ist schon paradox, dass, obwohl Algorithmen heute mehr denn je unseren Alltag und das gesellschaftliche Miteinander prägen, sie für viele von uns unsichtbar sind. Nicht zuletzt daher sind sie vielerorts Gegenstand öffentlicher Debatten. Bestrebungen, die Strukturen und Wirkweisen zu erklären und zur Transparenz von algorithmischen Systemen beizutragen, sind heute zur gesellschaftlichen Aufgabe geworden – eine Aufgabe, die zunächst der Informatik vorbehalten war, wie wir mit dem Eingangszitat verdeutlichen wollen. Im Jahr 1962 begann der US-amerikanische Informatiker Donald E. Knuth sich mit den Grundlagen von Algorithmen und Datenstrukturen zu beschäftigen. 1968 erschien „The Art of Computer Programming" als mehrbändige Publikation. Sie zählt zu den Standardwerken der Informatik und bildet damit eine wichtige Grundlage für das Verständnis über rechenbasierte Architekturen[1].

1 Knuth, der von der Kunst der Programmierung spricht und dazu auffordert Algorithmen zu hinterfragen, bietet uns heute einige Anknüpfungspunkte, um Algorithmen auch in interdisziplinären Kontexten zu thematisieren. Es ist weniger die Art und Vielfalt der Algorithmen, die Knuth für seine Zielgruppe auch exemplarisch verhandelt, sondern vielmehr wie er einerseits das Verhältnis von Mensch und Maschine in seinem Schreiben und Denken über Algorithmen bestimmt und andererseits sein Handeln ihn selbst definiert. Als er die überarbeitete Neuauflage des zweiten Bandes verfasste, forderten Knuth Unzulänglichkeiten der damaligen Satztechniken heraus, woraufhin er schließlich mit TeX sein eigenes digitales Satzsystem entwickelte, um die mathematischen Problemstellungen und Erkenntnisse zu Algorithmen angemessen darstellen zu können. Während TeX heute als Standardwerkzeug für Publikationen in der Mathematik und den Naturwissenschaften gilt, ist vor allem der Entstehungsprozesse

Algorithmen schlagen uns heute neue Freundinnen und Freunde vor, geben uns Buchempfehlungen. Sie beeinflussen unseren Musikgeschmack (Goldschmitt und Seaver 2019) und sie sind die Grundlage für die Berechnung unserer Kreditwürdigkeit (O'Neil 2016, Eubanks 2018). Daneben organisieren sie ganze Arbeitsabläufe (Rosenblat 2018) und unterstützen Ärztinnen und Ärzte, um schnellere Diagnosen zu erstellen (Roy et al. 2020). Trotz ihrer vielfältigen Einsatzszenarien und Wirkweisen entziehen sie sich nicht selten unserem Verständnis und genau dadurch entfalten sie eine besondere Faszinationskraft. Algorithmen sind nicht sichtbar und trotzdem allgegenwärtig und spürbar. Besonders dann, wenn fehlerhafte Berechnungen direkte Auswirkungen auf das Leben der Menschen haben. In ihrem Band „Hello World" skizziert Fry (2019) sehr anschaulich eine von Algorithmen geprägte Welt, die eben mit dieser Grußformel startete und eine expandierende Auswirkung auf unsere Leben hat: Über den Einfluss von Daten, die Auswirkungen auf Bereiche wie Verkehrsinfrastruktur, Mobilität, Medizin, Justiz und Kriminalität, zu Kunst und der eigentlichen Frage, wer tatsächlich mächtig ist. Zwischen Skepsis und Vertrauen beziehungsweise Zutrauen in die neue Technik zeigt die Autorin die Grenzen von scheinbar gerechten und objektiven algorithmischen Entscheidungsprozessen auf. Einige dieser Beispiele, die hier zur Problematisierung der gesellschaftlichen, kulturellen und politischen Herausforderungen beitragen sollen, haben ihren Ursprung in den USA oder Großbritannien.

In Deutschland verlaufen viele gesellschaftliche Entwicklungen jedoch anders als in den USA oder auch Japan, da automatisierte Systeme hierzulande unter anderen Voraussetzungen operieren, wenn sie überhaupt eingesetzt werden[2]. Es wäre daher zu kurz gegriffen, würde man die internationalen Debatten über die Einflüsse von Algorithmen einfach auf Deutschland übertragen. Dennoch ist das Wissen über automatisierte Systeme und ihre teils impliziten Auswirkungen auf gesellschaftliche Prozesse und Aushandlungen eine gemeinsame Triebfeder des gesteigerten Interesses an Algorithmen.

Unabhängig davon, ob und wie durch Algorithmen soziale Ungleichheiten reproduziert werden, macht schon ihre Existenz an sich eine gesellschaftliche Schieflage deutlich, denn das Wissen über Algorithmen ist nach wie vor in weiten Teilen der Gesellschaft diffus, weshalb man nach wie vor auch von Black Boxes sprechen kann (Pasquale 2015, Zweig 2019). Algorithmen zu verstehen, wird und ist also sowohl für jede und jeden Einzelnen eine Her-

hinsichtlich des Empowerments im Umgang mit digitalen Technologien bemerkenswert und ein geeigneter Anlass, um in die Komplexität von Selbstbestimmung und Autonomie im digitalen Zeitalter einzuleiten.

2 Einen Einblick in die emotionalen und gesellschaftlichen Einflüsse humanoider Robotik versucht der vielfach ausgezeichnete Film „Hi, AI" aus dem Jahr 2019 zu geben (https://hiai-film.de/ [Zugriff:18.08.2021]). In dokumentarischen Episoden werden hier technologische Entwicklungen vorgestellt und in reale, teilweise emotional sehr bewegende, Handlungsräume in Japan und den USA eingebettet.

ausforderung. Es ist daher auch Aufgabe wissenschaftlicher Auseinandersetzungen, sich den daraus hervorgehenden Problemstellungen zu widmen und sowohl empirisch-analytisch als auch theoretisch-konzeptionell danach zu fragen, wie Autonomie in einer von unsichtbaren Algorithmen durchzogenen Welt gelingen kann. Dieser Frage haben sich die Autorinnen und Autoren des Bandes durch ihre unterschiedlichen disziplinären Zugriffe angenommen. Der Band versammelt Beiträge aus Erziehungs-, Medien- und Politikwissenschaft, Soziologie, Informatik und Philosophie.

Algorithmen und Autonomie: Bezugspunkte eines komplexen Verhältnisses

Das Verhältnis von Algorithmen und Autonomie wird im vorliegenden Band im Zusammenhang von Selbstbestimmung und Datenpraktiken verhandelt. Es handelt sich (1) um das Verhältnis von Selbstbestimmung, Autonomie und Freiheit, (2) um Bildung, Algorithmen und Datenpraktiken sowie (3) um Algorithmen, Robotik und ethische Fragen.

Man kann das komplexe Verhältnis von Algorithmen und Autonomie aus einer Vielzahl von Perspektiven heraus beleuchten und genau dadurch entfaltet es eine hohe Faszinationskraft. Die Formulierung dieser drei Aspekte ist das Ergebnis des interdisziplinären Austauschs sowohl im Veranstaltungsformat der Ringvorlesung, die im Sommer 2020 an der Otto-von-Guericke-Universität in Magdeburg von den Herausgebenden dieses Bandes organisiert wurde, als auch daran anschließend. Dabei werden verschiedene Theorielinien aufeinander bezogen. Eine Skizzierung der Grundlinien, die dieses komplexe Verhältnis ausmachen, halten wir daher aus zwei Gründen für angebracht. Erstens dient die Einordnung dieser Bezugspunkte der thematischen und konzeptionellen Selbstverortung. Es geht uns darum, zu erläutern, welche Akzente und Anknüpfungspunkte sich für die unterschiedlichen disziplinären Perspektiven ergeben. Zweitens reicht die Bestimmung der hier genannten Bezugspunkte über disziplinäre Grenzen hinaus und soll dazu beitragen, den Lesenden Einsichten zu geben und individuelle Anschlüsse herzustellen. Es werden über die drei Bezugspunkte Leerstellen gebündelt und identifiziert, die in den einzelnen Beiträgen dezidiert aufgegriffen werden. Wir wollen diese Aspekte ausgehend von den disziplinären Zugriffen konturieren, um so eine vielfältige und für den Band spezifische Perspektive zu entwickeln. Dafür wird zunächst der Autonomiebegriff aus einer philosophischen Perspektive betrachtet, um so die Grenzen und Reichweiten von Autonomie in einer algorithmisierten Welt grundlegend bestimmen zu können. Wie sich Algorithmen im Zusammenhang von Datenpraktiken verorten lassen und welche Implikationen sich hieraus für das Wissen über diese Strukturen und Bildung ergeben, wird über die Forschung aus einer erziehungswissenschaftlichen Perspektivsetzung herausge-

arbeitet. Abschließend wird aus einer informatischen Perspektive herausgearbeitet, wie sich Autonomie im Schnittfeld von Automatisierung und Ethik bewegt.

Autonomie und ihre Möglichkeit in einer algorithmisierten Welt

Freiheit ist der zentrale Pfeiler des Selbstverständnisses westlicher liberaler Gesellschaften. Frei zu sein, bedeutet für uns ein selbstbestimmtes, d. h. autonomes Leben zu führen. Der Anspruch auf Autonomie ist gegenwärtig eine Selbstverständlichkeit, sei es auf politischer oder individueller Ebene: Jeder will selbstbestimmt agieren und mitbestimmen können, die eigenen Geschicke lenken und nicht gelenkt werden.

Begriffsgeschichtlich liegen die Wurzeln der Autonomie im Griechischen: Sie bedeutet Selbstgesetzgebung (auto = selbst, nomos = Gesetz). Das Postulat der Autonomie des Menschen findet im Zuge der europäischen Aufklärung Eingang in die westlichen Gesellschaften. Zwar findet sich der Autonomie-Begriff auch in der Antike, doch vorrangig als politische Kategorie im Sinne der Unabhängigkeit der griechischen Stadtstaaten von äußerer Einmischung. Die Begriffswende zum Individuum hin vollzieht sich im aufklärerischen Europa der Neuzeit (vgl. Quante 2011): Dem Menschen wird die grundlegende Fähigkeit zur Selbstbestimmung zugesprochen, verbunden mit dem Postulat, aus Bevormundung und Fremdbestimmung herauszutreten, kraft der ihm verliehenen Vernunft: Sapere aude! Was in der Aufklärung folglich grundgelegt wird, ist ein Bild vom Menschen als eines vernunft- und daher freiheitsfähigen Wesens – eine Prämisse die auch heute noch, implizit oder explizit, Geltung hat und erst individuelle sowie politische Selbstbestimmung und Gestaltung möglich macht.

Mag das aufklärerische Menschenbild auch immer noch prägend für die westlichen Gesellschaften sein, so ist nicht zu übersehen, dass sich die Vorstellung des Menschen als rationalem und autonomem Subjekt zahlreichen Angriffen ausgesetzt sah und sieht. Die Geschichte des im Kern vernunft- und selbstbestimmungsfähigen Subjekts ist auch die seiner immerwährenden Widerlegung: Die marxistische Gesellschaftstheorie ließ das Subjekt in den ökonomisch-materialistischen Umständen aufgehen (Marx 2005); die Psychoanalyse verwies auf das Triebleben und das Unbewusste, durch die der Mensch gesteuert wird (Freud 2013); aktuell lehren uns die Neurowissenschaften, das autonome Subjekt sei maßgeblich determiniert von neuronalen Vorgängen, seine Freiheit nur Illusion (Roth 1996). Zu diesen Beispielen, die man beliebig erweitern könnte, fügt sich die poststrukturalistische Philosophie, die das Ende des Menschen einzuläuten beansprucht, indem sie in ihm nurmehr das Produkt diskursiver Hegemonie erkennt: Das Subjekt ist hier nicht Ur-

sprung seiner selbst, sondern Effekt der herrschenden Machtstrukturen. Die Geschichte des autonomen Subjekts scheint auserzählt (Foucault 1997).

Die Spannung zwischen immerwährenden Versuchen philosophischer und wissenschaftlicher Widerlegung des autonomen Subjekts und der lebenweltlich-praktischen Voraussetzung eines solchen setzt sich im digitalen Zeitalter fort. Die Durchherrschung unseres Lebens durch das Digitale fordert den Anspruch auf Autonomie in einem weitreichenden Maße heraus, gerade da die Vorteile, die die digitale Vernetzung mit sich bringt, die Ambivalenzen einer durchalgorithmisierten Wirklichkeit nur allzu leicht übersehen und übergehen lassen. Derer aber sind viele; sie äußern sich in Hinblick auf die Selbstbestimmung auf diversen Ebenen:

– Die zahlreichen Erleichterungen, die intelligente Systeme dem Menschen verschaffen, ziehen Kreise der Entmündigung, denn wer sich auf algorithmische Systeme verlässt, Aufgaben an diese delegiert, verliert an eigenen Fähigkeiten, die nur mittels mühsamer Selbsttätigkeit erhalten und gefördert werden können. So mag manches schneller und unkomplizierter gelingen, doch zum Preis einer stetig wachsenden Abhängigkeit von digitaler Technik. Auch können Empfehlungen, durch Algorithmen auf das eigene Profil zugeschnitten – sei es in Hinblick auf Musik, Serien, Zeitungsartikel oder andere Konsumartikel – als bereichernd und den Alltag entlastend empfunden werden; gleichzeitig aber fördern sie die Tendenz, im Eigenen zu verbleiben, statt dieses zu hinterfragen und Neues zu entdecken. „Echokammern“ und „Filterblasen“ sind die gegenwärtigen Stichworte für eine so gelagerte Entwicklung der politischen Diskurse (Pariser 2011, Nguyen 2020). Doch auch auf viel grundlegenderer Ebene zeitigt dies problematische Auswirkungen, setzt doch Selbstbestimmung – ein autonomes Verhältnis zu sich und zur Welt – eine kritische, also mit dem Alltäglichen und Selbstverständlichen brechende und diese hinterfragende Haltung voraus.
– Die Digitalisierung macht den Menschen sukzessive zu einer berechenbaren Größe: Er wird zusehends auf ein Gemenge von Daten und Informationen reduziert. Exemplarisch lässt sich dies am Self-Tracking im Rahmen der Quantified-Self-Bewegung verdeutlichen (Crawford et al. 2015, Duttweiler et al., 2016, Damberger 2021a). Mittels verschiedener Apps, Sensoren-Armbändern, Smart-Watches etc. werden Daten zu Fitness, Ernährung, Bewegung, Schlafrhythmus etc. zusammengetragen, um aus dieser Ansammlung von Datenpunkten die Errechnung von Selbstoptimierung zu ermöglichen. Dabei wird nichts Geringeres beansprucht als die Erfüllung des wohl grundlegendsten Strebens der Philosophie seit ihren Anfängen: Selbsterkenntnis – so als gehe menschliches Leben gänzlich in algorithmischer Verfasstheit auf. Dies kritisch zu prüfen und zu hinterfragen, ist Aufgabe der philosophischen Anthropo-

logie sowie der Phänomenologie (Fuchs 2020). In Hinblick auf die Frage der Autonomie tut sich ferner eine fundamentale Problematik auf, denn: Wer berechenbar ist, ist beherrschbar, steuerbar und in letzter Instanz manipulierbar, muss doch eingeräumt werden, dass unsichtbar bleibt, wer auf die eigenen vitalen Daten Zugriff hat und zu welchem Zwecke diese verwendet werden.

Damit sind exemplarisch einige Ebenen hervorgehoben, die philosophische Problemstellungen berühren. Eine zentrale Herausforderung für die Philosophie ist schließlich die Übertragung des Begriffs der Autonomie auf intelligente Systeme – man spricht von autonomem Fahren, von autonomen Waffensystemen oder gar von autonomen KI-Künstler*innen. Aus der Bedeutung der Autonomie als Selbstbestimmung wird eine Form der Selbsttätigkeit. Problematisch ist dies, insofern die Begriffsverschiebung zugleich eine Zuschreibung von zentralen menschlichen Qualitäten suggeriert: Autonomie umfasst die Ebenen von Reflexion, Entscheidung, Handlung und Verantwortung, Akte, die von intelligenten Systemen, wenn überhaupt, nur simuliert, nicht aber verkörpert werden können. Ihnen fehlt nebst einem Selbst- und Weltverhältnis – dem Vermögen, sich selbst und die Welt zu vergegenständlichen – die leibliche und existentielle Teilhabe an der Wirklichkeit.

Bildung, Algorithmen und Datenpraktiken

Wenn wir heute über Bildung sprechen, dann können wir das nicht mehr ohne den Einfluss von digitalen Medien und ihren Technologien machen. In der Konsequenz ist *Bildung in der digitalen Welt* eine komplexe und komplizierte Aufgabe. Daran schließt das gleichnamige Strategiepapier der Kultusministerkonferenz (KMK) aus 2016 an.

Die KMK-Strategie stellt ein umfassendes Handlungskonzept zur digitalen Entwicklung des Bildungswesens innerhalb von Deutschland dar. Eine wesentliche Zielsetzung der KMK-Strategie ist die Förderung von digitalen Kompetenzen bei Lehrkräften, damit diese schließlich den Schüler*innen eine selbstbestimmte Teilhabe in der digitalen Welt ermöglichen können (KMK 2016). Es geht dabei also um die Aus- und Weiterbildung von Lehrenden und die Entwicklung von Lehrkonzepten in Hinblick auf digitale Kompetenzen und den gezielten Einsatz von Bildungsmedien und andererseits den Auf- und Ausbau der technisch-digitalen Infrastruktur der Schulen. Die digitale Transformation zeichnet ein hochdynamisches Bild und hat auch die KMK zur Weiterentwicklung des Strategiepapiers bewogen, welches 2021 veröffentlicht wurde. Hier zeigt sich eine Hinwendung zur Digitalität. So wird die Kultur der Digitalität (Stalder 2016) in Debatten über eine zeitgemäße Schulkultur überführt, um für das Leben in der digitalen Welt fit zu machen. Neben dieser

bildungspolitischen Perspektive wird Bildung schon länger unter Berücksichtigung angrenzender Felder verhandelt und die Konstitution dessen, was der Mensch ist, im Verhältnis zu den äußeren und zeitgemäßen Rahmenbedingungen gesehen. Dazu zählt auch die Reflexion darüber, was Digitalität eigentlich ausmacht (Hauck-Thum et al. 2021). Daneben finden kritische Reflexionen auch über ideologische Aspekte der Digitalisierung im Zusammenhang zur Bildung statt, die sich einerseits an der Kritischen Pädagogik (Damberger 2021b) orientieren und andererseits bildungspolitische Herausforderungen und Problemlagen aus einer medienpädagogischen Position heraus adressieren (Dander 2018). Die Zielsetzungen der KMK spiegeln sich auch in verschiedenen interdisziplinären Forschungsarbeiten wieder, die sich in Perspektiven auf *datengetriebene Schulen* beziehen und damit im Schnittfeld von Medienpädagogik und Informatik interdisziplinäre Perspektiven und Entwicklungen im Schulkontext herausheben (Schiefner-Rohs et al. 2021).

Das digitale Zeitalter ist geprägt von Berechnungen, Daten und Vernetzung und auch wenn die Vermessung des Menschen keine neue Idee ist, hat sie mit dem digitalen Wandel eine neue Dimension erreicht. Sei es die individuelle Selbstvermessung, die Überwachung der Leistungen in der Arbeitswelt oder Bestrebungen zur Verbesserung von Lernerfolg. Learning Analytics und Big Data sollen dies sicherstellen. Sie werden aus erziehungswissenschaftlicher Sicht dahingehend befragt, inwiefern sie als Instrumente einer neuen kybernetischen Steuerung an Schulen gesehen werden können (Reinisch 2020). Die optimierte Gestaltung von individuellen Lernerfahrungen und die effektive Sicherstellung des Lernerfolgs ist dabei jedoch von einer Rhetorik geprägt, die an die Ökonomisierung der Bildung erinnert und damit nur wenig mit der freien Entfaltung des Menschen zu tun hat. Dataveillance und Learning Analytics werden damit als Herausforderung erziehungswissenschaftlicher Reflexionen gesehen, die dann auch im Lichte von Technikfolgeabschätzungen diskutiert werden (Karcher 2020). In dem von Christian Leineweber und Claudia De Witt herausgegebenen Online Sammelband zu „Algorithmisierung und Autonomie im Diskurs“ werden fortwährend grundlegende Reflexionen versammelt, die den Einfluss von algorithmischen Strukturen aufarbeiten und deren Bedeutung sowie Folgen für die Autonomie des Menschen aus einer erziehungswissenschaftlichen Perspektive diskutieren. Dabei geht es unter anderem um Fragen zur Datenkritik und Urteilsfähigkeit (Leineweber 2021) oder auch Autonomieverhältnissen durch Recommender-Systeme in der Hochschullehre (Waldmann & Wunder 2021).

Mit dieser groben Konturierung, die keinerlei Anspruch auf Vollständigkeit erhebt, wird jedoch schon deutlich, dass Bildung in der digitalen Welt weit mehr als nur formale Bildung ist, denn digitale Technologien wirken sich in unterschiedlichen Lebenslagen auf die Konstitution der Selbst- und Weltverhältnisse von Menschen aus. Nach Jörissen und Marotzki (2009) finden soziale Aushandlungsprozesse mehr und mehr auch in den unterschiedlichen sozialen

Arenen statt, die erst durch die digitalen Medien eröffnet werden. Die Autoren beschreiben einen strukturtheoretischen Ansatz von Medienbildung, der individuelle Konfigurationen und Transformationen von Selbst- und Weltverhältnissen im Zusammenhang zu gesellschaftlichen Veränderungsprozessen verhandelt und damit Medialität im Schnittfeld von Medien- und Bildungstheorie erfasst. In diesem Schnittfeld werden die Strukturen der digitale Medien auch empirisch dahingehend untersucht, welche Lern- und Bildungspotenziale ihnen eingeschrieben sind. Aus einer erziehungswissenschaftlichen Sicht sind hier nicht nur Lern- und Bildungspotenziale relevant, sondern auch deren Entfaltung und die sich vollziehenden Bildungs- und Subjektivierungsprozesse. Die sozialen Arenen sind damit auch von Macht- und Herrschaftsstrukturen geprägt, um Spielräume von Freiheit und Kontrolle (Biermann & Verständig 2017) zu adressieren oder ganz grundlegend das Verhältnis von Digitalisierung, Subjekt und Bildung (Dander et al. 2020) zu bestimmen.

Medienbildung wird damit direkt ins Verhältnis zur informatischen Bildung gesetzt (Rummler et al. 2016), was wiederum die Tür zu digitalen Infrastrukturen, also Algorithmen und Dateninfrastrukturen öffnet. In diesem Moment wird in zweifacher Hinsicht von Algorithmen gesprochen: *Erstens* haben Algorithmen eine lange Geschichte und verweisen auf eine noch längere mathematische Tradition. Sie sind in ihrer Beschaffenheit und Klassifikation mit fortschreitender Rechenleistung in ihrer Komplexität avanciert. Wenn heute von Algorithmen die Rede ist, dann geht es meist um datengetriebene Verfahren. In einer Welt, in der immer mehr und immer Daten ganz unterschiedlicher Qualität zur Verfügung stehen, werden diese Verfahren zu einem integralen Bestandteil in der Optimierung und im Design von Abläufen. Die Berechnung und Auswertung von großen Datenbeständen durch Computer hat den großen Vorteil, dass sie entgegen menschlicher Interpretation nicht zu einer Überinterpretation einiger weniger und Vernachlässigung vieler anderer Daten neigen. Sie haben zugleich den Nachteil, dass menschliche Entscheidungen vom Design und der Funktionalität dieser Systeme abhängig sind. *Zweitens* meinen wir auch grundlegend die Interaktion von Mensch und Maschine und beziehen die Praktiken ein, die sich im Zusammenspiel ergeben. Nach Reckwitz (2003) sind dies wiederkehrende Handlungen, die als Ausdrucksformen von sozialen Ordnungen verstanden werden können (vgl. ebd., S. 289 f.). Daher halten wir den Begriff der Datenpraktiken für passend, um die Komplexität von algorithmischen Systemen über die technologischen Zusammenhänge hinaus zu erfassen. Praktiken im Umgang mit Algorithmen werden in informellen Lern- und Bildungskontexten weitestgehend über das Spiel mit Software und die Digitalen Medien thematisiert. Beispielsweise, wenn man sich in den algorithmischen Architekturen bewegt, sich durch bestimmte Anfragen Informationen verschafft und diese dann bewertet. Dadurch soll der Blick nicht nur auf die uns anstrahlenden Displays gelenkt werden, sondern auch auf die da-

hinterliegenden Prinzipien und Herausforderungen im Umgang mit algorithmischen Systemen.

Algorithmen, Robotik und ethische Fragen

Die Informatik nimmt bei den zuvor skizzierten Entwicklungen eine besondere Rolle ein, die heute weit über den Bau und die Programmierung von reinen Rechenmaschinen hinausgeht und damit in weite Teile der Gesellschaft hineinreicht. Im Mittelpunkt steht dabei die Information, und die Grundlage bildet die universale Logik der Algorithmen und Berechnung, sei es die Wetter- und Klimavorhersage und das allgemeine Zirkulationsmodell, die Signalverarbeitung, die durch die schnelle Fourier-Transformation realisiert wird oder intelligente Assistenzsysteme.

Aus informatischer Sicht ergeben sich aus den zuvor skizzierten Entwicklungen und Phänomenen auch erst neue Handlungsräume. Es entstehen Forschungsgebiete, die sich mit Fragen der Erklärbarkeit von automatisierten Entscheidungsprozessen durch intelligente Systeme beschäftigen, um somit Vertrauen in solche Systeme (zurück) zu gewinnen oder Forschenden und Entwickelnden solcher Systeme zu ermöglichen, ihre automatisiert erstellten Algorithmen schlussendlich selbst wieder zu verstehen und basierend auf diesen Erkenntnissen Potenziale für die weitere Optimierung zu schaffen. Ethische Betrachtungen haben dabei eine – für das junge Forschungsgebiet der Informatik – lange Tradition und viele kritische Diskurse finden ihren Ursprung in der Science Fiction Kurzgeschichte „Runaround“, die 1950 im Sammelband „I, Robot“ des Autors Isaac Asimov erschien. Hier wurden erstmals ethisch motivierte Grundregeln definiert, die ein Roboter bei der Interaktion mit Menschen einhalten muss: Die drei Regeln der Robotik[3], die vermutlich jede und jeder Robotikforschende gleich zu Beginn der Ausbildung kennengelernt hat. Mittlerweile haben viele Organisationen eigene Standards und Richtlinien definiert, die versuchen die technische Entwicklung sowohl durch rechtliche Richtlinien als auch durch Empfehlungen für die Entwicklerinnen und Entwickler intelligenter Technologien so zu beeinflussen, dass die gesellschaftlich gegenwärtig akzeptierten ethischen Leitlinien eingehalten werden[4]. Neben of-

3 In der Kurzgeschichte werden die folgenden Regeln definiert: „1. A robot may not injure a human being or, through inaction, allow a human being to come to harm. 2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law. 3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.“

4 Siehe z.B. IEEE P7000 – Model Process for Addressing Ethical Concerns During System Design (https://standards.ieee.org/project/7000.html) und die etwas breiter aufgestellte Initiative „Ethics in Action“ (https://ethicsinaction.ieee.org/), sowie die „Ethik-Leitlinien für eine vertrauenswürdige KI“ der EU (https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai) und die OECD AI Richtlinien (https://oecd.ai/en/ai-principles), sie sich

fiziellen etablierten Organisationen finden sich auch immer wieder kritische Wissenschaftler in interdisziplinären Workshops zusammen, die aktuelle Entwicklungen der Digitalisierung hinterfragen und teilweise in Interdisziplinär getragenen Initiativen münden, wie bspw. die „The Digital Humanism Initiative“[5], die 2019 das „Vienna Manifesto on Digital Humanism“ veröffentlicht hat und einen kritischen Diskurs zu digitalen Technologien einfordert.

Zu den einzelnen Beiträgen des Bandes

Die Beiträge des Bandes thematisieren die zuvor skizzierten Problemstellungen und Phänomene auf Algorithmen und Autonomie im digitalen Zeitalter aus unterschiedlichen Perspektiven. Dabei werden disziplinäre Einsichten aus der Erziehungswissenschaft, Informatik, Philosophie mit Beiträgen aus Politik- und Kommunikations- und Medienwissenschaft ergänzt. Der Band versammelt einerseits Beiträge, die auch schon als Vortragsimpulse im Rahmen der interdisziplinären Ringvorlesung „Autonomie im digitalen Zeitalter?! Suchbewegungen zwischen Freiheit und Kontrolle“ im Sommersemester 2020 an der Otto-von-Guericke-Universität in Magdeburg thematisiert und auch über einen hierzu produzierten Podcast veröffentlicht wurden. Andererseits wurden gezielt Beitragende aus unterschiedlichen Disziplinen um ihre Perspektiven angefragt, um ein möglichst heterogenes und breites Spektrum an Ansichten zu diesem Themenkomplex zu gewinnen. Alle Autorinnen und Autoren haben sich der Frage angenommen, wie das Verhältnis von Algorithmen und Autonomie aus ihrer Disziplin heraus zu denken ist und welche Herausforderungen mit den jeweiligen Auseinandersetzungen einhergehen.

Im Schnittfeld von Medien- und Bildungstheorie konturieren Janne Stricker und Dan Verständig Freiheit im digitalen Zeitalter. Sie erfassen den Freiheitsbegriff in seinen unterschiedlichen Facetten vor allem hinsichtlich der individuellen Selbstbestimmung. Dabei stellt die individuelle Herstellung von Orientierung einen wichtigen bildungstheoretischen Bezugspunkt dar. Prozesse, der Berechnung und Automatisierung erlauben, so die These, nicht nur Bestimmtheit, sie eröffnen vielmehr auch Unbestimmtheitsräume, zu denen sich die Menschen verhalten müssen. Sie sind gewissermaßen dazu verpflichtet, sich selbstbestimmt zu verhalten. Diese Figuration wird dann durch verschiedene Paradoxien um Individualität, Differenzierung und Diskriminierung herausgestellt. Dabei wird eine Argumentation entfaltet, die Algorithmen und Daten eine performative Kraft zuschreibt. Algorithmen repräsentieren nicht nur die Welt, in der wir leben, sie prägen unsere Welt und stehen im engen Wechselverhältnis zu menschlichen Handlungen und Entscheidungen. Daten

sowohl an Wissenschaftler, aber auch die Gesetzgeber der Mitgliedsländer richten, von denen sie 2019 erstmals unterzeichnet wurden.

5 https://dighum.ec.tuwien.ac.at/

sind dann nicht das objektive Resultat von vermeintlich neutralen Berechnungen, sie sind von sozialen sowie kulturellen Einflüssen geprägt und damit keineswegs neutral.

Aus einer bildungs- und technikphilosophischen Perspektive heraus zeigt Thomas Damberger in seinem Beitrag analytisch auf, wie der Geist in die Maschine gelangt ist und damit das Verhältnis von Mensch und Maschine im Digitalzeitalter prägt. Damberger skizziert, entlang der Geschichte der Maschinen, welche Möglichkeiten und Herausforderungen sich durch kybernetische Systeme für das Menschsein ergeben. Ziel eines jeden kybernetischen Systems sei die Erreichung und der Erhalt der Homöostase, also eines Zustands des Gleichgewichts zwischen dem, was sein soll, und dem, was ist. Wie das im Zusammenhang mit dem Geist in der Maschine zu verstehen ist, leitet Damberger durch die Entgrenzung der Maschine auf drei Ebenen her: Auf der ersten Ebene entgrenzen wir uns durch das Selftracking. Unser quantifiziertes Selbst (*Quantified Self*) erschaffen wir, indem wir mithilfe von Tools und Technik immer mehr (Körper-)Daten sammeln und auswerten. Die zweite Ebene der Entgrenzung findet dann statt, wenn der menschliche Organismus kybernetisiert wird, sprich wenn wir zum *Cyborg* werden. Damberger diskutiert dabei, ob und inwiefern die Technik sogar ein (besserer) Teil unseres Körpers werde und sogar mit dem Gehirn kommunizieren könne. Als dritte und somit letzte Ebene der Entgrenzung stellt der Autor den Gedanken des *Mind Uploadings* vor, also die Externalisierung und Übertragung des menschlichen Bewusstseins auf einen Computer. In dieser Entgrenzung steckt ebenfalls eine Eingrenzung von Freiheit und Kontrolle, wie Damberger in Hinblick auf die Entwicklung von Mensch-Maschine sowie Subjekt und Welt zeigt, indem er die Diskrepanz zwischen dem, was sein soll, und dem was ist, kritisch diskutiert.

Estella Ferraro und Friedrich Wolf behandeln in ihrem Beitrag die Frage, wie KI im Film dargestellt wird und wie dies im Zusammenhang zur Darstellung des Menschen verhandelt wird. Ausgehend von einer Perspektive im Schnittfeld von Erziehungs- und Medienwissenschaft stecken sie entlang einer filmethischen Analyse am Beispiel des Films Ex Machina (2015) die Grenzen und Reichweiten menschlicher Autonomie ab. Die Verflechtungen von Körperlichkeit, Autonomie und Technologie sind zentrale Motive in erziehungs- und sozialwissenschaftlichen Diskursen. Die filmethische Analyse bietet einen Zugang, um das Verständnis über diese Grundbegriffe und ihre Reichweiten im Horizont von KI-Systemen erneut zu bestimmen. Die Analyse wird in Anlehnung an Turing geleitet von drei Fragen: Wer wird imitiert? Wie wird imitiert? Wozu wird imitiert? Damit eröffnet die reflexive Auseinandersetzung über das audiovisuelle Medium Film einen weiteren Einblick in das tiefere Verständnis und die Rhetorik über digitale Technologien im Allgemeinen und KI im Besonderen. Damit eignet sich das Medium Film grundlegend, um gesellschaftsrelevante Diagnosen zu stellen und im Hinblick auf die Medienbildung zu beziehen. Der Film Ex Machina eignet sich dabei besonders gut, um Fragen

von Macht, Gender und Materialisierung von KI kritisch zu hinterfragen und entlang der Referenz von Autonomie zu diskutieren.

Katharina Simbeck widmet sich in ihrem Beitrag ebenfalls dem Schnittfeld von Bildung und KI, legt dabei aber den Fokus auf Fairness und Diskriminierung durch *Machine Learning* im Kontext von Learning Analytics. Durch die Verwendung von digitalen Lernmaterialien und Lernmanagementsystemen entstehen über verschiedene Datenpraktiken immer mehr Daten über Lernende, Lehrende und Lernprozesse, welche schließlich ausgewertet und für unterschiedliche Zwecke genutzt werden können. Hieran schließt nicht nur ein bestimmtes Bildungsverständnis an, sondern auch die grundlegende Frage, inwiefern digitale Technologien die Beziehung von Lehrenden und Lernenden untereinander und zum Lernen überhaupt beeinflussen. Es geht damit um die Implikationen von Learning Analytics und die adäquate Erfassung von Lernerfolgen. Dabei ist es, so Simbeck, weder nötig noch möglich, dass Lehrende und Lernende im Bildungsbereich eingesetzte KI-Systeme im Detail verstehen, nachvollziehen oder überprüfen können, da die Systeme große Datenmengen, die sich kontinuierlich ändern verwenden und die dahinter liegenden Modelle und Algorithmen auch für Expert*innen nur mit großem Zeitaufwand nachvollziehbar sind. Für viele Anwendungsgebiete ist es jedoch notwendig, die Ausgaben des Systems, also die Einflussfaktoren auf den Output verstehen zu können, um diese zu akzeptieren oder gegebenenfalls auch korrigieren zu können. Wenngleich Lehrende, Lernende und Entscheidungsträger*innen im Bildungssystem nicht über umfangreiches Spezial- und Expert*innenwissen der Technologien und Methoden verfügen müssen, sollten sie sich jedoch bewusst sein, dass den Chancen, die sich durch datengetriebene Verfahren ergeben, eine Reihe von Herausforderungen gegenüberstehen. Diese Herausforderungen betreffen zum einen Datenschutz, Transparenz und die informierte Möglichkeit, auf die Datenverarbeitung zu verzichten. Dementsprechend ergeben sich auch Herausforderungen hinsichtlich der Transparenz und Diskriminierungsfreiheit der eingesetzten Systeme, die sich entlang der Fragen abbilden, ob oder inwieweit die Daten hinsichtlich verschiedener Merkmale, wie Gender, sozialen Gruppen oder beispielsweise auch Migrationshintergrund unverzerrt sind.

Die Problematik der *gleichberechtigten Autonomie* wird von Andrea Wolffram aufgegriffen. Dabei geht es der Autorin um gesellschaftliche Zusammenhänge aus einer techniksoziologischen Perspektive mit Schwerpunkt auf Gender Studies. Basierend auf dem dritten Gutachten des Gleichstellungsberichts der Bundesregierung 2021 werden unterschiedliche Gesichtspunkte einer Digitalisierung aus Geschlechterperspektive im Schnittfeld von Diskriminierung und Ungleicheit aufgezeigt. Ziel ihres Beitrags ist es, bestehende Analysen und Phänomene der Ungleichheit und Geschlechterdiskriminierung vor dem Hintergrund der Digitalisierung aufzuzeigen. Wolffram skizziert dabei die vier Stränge der Kritik an der Autonomie und übersetzt diese

in Hinblick auf Digitalisierung. Es geht ihr 1) um Mechanismen der Ausgrenzung und Ausblendung aus digitaler Öffentlichkeit sowie lebensweltlichen Zusammenhängen, 2) um die Ungleichstellung zwischen den Geschlechtern in digitalen (Erwerbsarbeits-)Welten, 3) um die Berücksichtigung digitaler Gewalterfahrung und Reflexion digitaler Gewalt als gesellschaftliches Struktur- und Herrschaftsmuster und 4) schließlich um die Herstellung gleichberechtigter Autonomie in digitalen Lebenswelten. In ihren Analysen wird deutlich, wie sich das Verhältnis von Macht und Geschlecht auf unterschiedlichen Ebenen konfiguriert. Eine Diagnose ihrer Untersuchung ist dabei, dass Frauen sowohl in der Technologieentwicklung als auch der -gestaltung sowie der Entscheidung über den technologischen Einsatz unterrepräsentiert sind. Das Thema der Ungleichheit wird dabei aus unterschiedlichen Perspektiven hergeleitet und exemplarisch problematisiert: Entlang der Ergebnisse der D21-Studie „Digital Gender Gap" werden von der Autorin die ungleichen Gestaltungsmöglichkeiten der Lebenswelt außerhalb professionalisierter Spezialbereiche sichtbar. Wolffram zeigt, dass durch bereits unterschiedliche Zugänge zu digitalen Technologien, die Möglichkeiten der gleichberechtigten Autonomie nicht chancengleich sind, sondern über verschiedene Freiheitsgrade soziale Ungleichheiten im digitalen Raum auch durch große Datenmengen reproduzieren. Big Data wird bei Wolffram weniger als technisches und vielmehr soziales Problem gefasst, bei dem vermeintliche Zuschreibungen von Objektivität bereits deutungsmächtig sind und somit Erkenntnisweisen sowie Wirklichkeitsaspekte an westlich-weiße Männlichkeit knüpft, wodurch andere Erkenntnisweisen drohen marginalisiert zu werden.

Christoph Sebastian Widdau und Carlos Zednik diskutieren in ihrem Beitrag aus einer philosophischen Perspektive den Begriff der Verantwortung am Gegenstand von KI. Der Ausgangspunkt der Überlegung beider Autoren ist der Mensch, der es schafft, etwas herzustellen, das selbstlernend, selbsttätig und autonom agieren kann. Menschen sind fähig opake KI-Systeme zu konstruieren und mit dieser „Selbstständigkeit" von Systemen, stellt sich die ethische Frage danach, wie mit Fehlern und den damit einhergehenden Folgen umgegangen wird. Am Topos der Opazität von KI-Systemen zeigen die Autoren beispielhaft auf, wie die Zurechnung und der Nachweis von Fehlfunktionen in Systemen der Künstlichen Intelligenz im Verhältnis zu Fragen der Verantwortung stehen. Opake Systeme der Künstlichen Intelligenz sind keine Naturgewalten, sondern regulierbare Systeme, sie sind nicht menschlich und somit auch keine Verantwortungsträger im eigentlichen Sinn, da die Systeme sich und ihr Verhalten nicht verantworten können. Aber wer trägt dann die Verantwortung? Die Nutzenden, die Programmierenden oder die Betreibenden der Anwendung? Oft können menschliche Programmier:innen selbst nicht vorhersagen, wie sich KI-Systeme verhalten, da es sich um opake Systeme handele. Damit rücken Fragen der Verantwortung in ganz unterschiedlichen Szenarien in den Mittelpunkt. Das Spektrum reicht vom Einsatz selbstfahrender Autos bis hin zu autonomen

Waffensystemen. Es handelt sich grundlegend um das Problem der Verantwortungslücke. Diese Problematik wird durch die Autoren entlang der Differenzierung von kausaler und moralischer Verantwortung auch exemplarisch erörtert und damit interdisziplinär anschlussfähig veranschaulicht. Widdau und Zednik plädieren in ihrem Beitrag schließlich dafür, die These der Verantwortungslücke im Zusammenhang mit KI-Systemen zurückzuweisen, denn es ginge nicht nur um Verantwortung und die Frage, ob man schuldig oder nicht schuldig sei, sondern vielmehr um den produktiven Umgang mit Fehlern und die Herausforderungen, den Umgang mit KI-Systemen und ihrer Opazität zu explizieren, um so schließlich die eigene moralische Verantwortung nicht zu negieren, sondern sich dazu ein produktives Verhältnis zu setzen.

Christina Kast widmet sich in ihrem Beitrag der Problematik zur sogenannten kreativen Künstlichen Intelligenz. Von einer philosophischen Warte aus wird erfragt, ob Algorithmen in einem menschlichen Sinne schöpferisch sein können. Von Interesse ist demnach, ob KI mehr als nur Hilfsmittel eines menschlichen Künstlers sein kann, nämlich selbst – dem Menschen gleich – autonomer Künstler. Grundintention des Beitrags ist es, den Beweis zu führen, dass dies zu verneinen ist, da Kunst als genuin menschliche Lebensäußerung, die maschinell weder erzeugt noch imitiert werden kann, zu verstehen ist. Der Künstler ist zunächst Mensch und darin, gemäß Helmuth Plessner, bestimmt von seiner exzentrischen Stellung in der Welt sowie seiner Leiblichkeit. Exzentrizität und Leib-Sein, so zeigt der Beitrag, sind genuine Voraussetzungen für die Möglichkeit von Verhältnis, Wahrnehmung und Erfahrung zur und von Welt, ohne die kein künstlerisches Schöpfertum möglich ist. Davon ausgehend wird dargelegt, dass der Mensch, welcher Künstler ist, sich wiederum durch ein spezifisches Weltverhältnis – sowie eine spezifische Wahrnehmung und Erfahrung von Welt – auszeichnet: Er sieht die Welt anders als die Anderen; dieses Anders-Sehen der Welt, das sich im Kunstwerk materialisiert, gründet zumeist in Formen eines negativen Weltbezugs: in Entfremdung, Leiden und Schmerz. Dem Algorithmus ist der Zugang zu dieser Welt der Negativität, wie auch zur exzentrischen wie leiblichen Existenz, versperrt.

Wolf Schünemann fokussiert in seinem Beitrag die politischen Ansätze der Regulierung von Internetinhalten und allgemeiner der Internetkontrolle in liberalen Demokratien. Das Hauptaugenmerk wird dabei auf die Entwicklung der Regulierungsdiskurse in Bezug auf digitale Medien anhand der darin erkennbaren legitimatorischen Strategien gelegt. Folgt man aktuell vorherrschenden Begründungsfiguren für die Regulierung von Online-Kommunikation in liberalen Demokratien, insbesondere im Hinblick auf die so genannte Desinformation, dann soll die gesteigerte Regulierungsaktivität in erster Linie dem Schutz gesellschaftlicher Autonomie und demokratischer (Informations-) Freiheit dienen. Doch inwiefern sind die derzeit diskutierten oder bereits implementierten Maßnahmen zur Regulierung von Inhalten, um Desinformationen einzudämmen beziehungsweise gänzlich zu unterbinden, tatsächlich ge-

eignet? Worin lassen sich die Grenzen und Reichweiten insbesondere dahingehend bestimmen, wie die verschiedenen Maßnahmen mit den grundlegenden Werten einer liberalen Demokratie übereinstimmen oder inwiefern sie womöglich selbst eine Gefahr für unsere freiheitliche Ordnung darstellen? Damit greift der Autor ein konfliktbehaftetes Spannungsverhältnis zwischen Meinungsfreiheit und informationeller Autonomie auf, welches gerade in der Tradition der deutschen staatsrechtlichen Betrachtung prominent diskutiert worden ist. Schünemann entfaltet seine Perspektive auf die Inhalteregulierung zunächst über die Rekonstruktion einiger historischer Linien, die sich über die Regulierung von Internetinhalten und das Ende des Exzeptionalismus beschreiben lassen. Daran schließt eine eine disziplinäre Positionierung an, die sich den medienpolitischen Entwicklungen und den damit einhergehenden Souveränitätsvorstellungen in unterschiedlichen Ausprägungen zuwendet, um so schließlich die Legitimation einer demokratischen Inhalteregulierung kritisch erfassen und ihre Reichweiten bestimmen zu können. Schließlich geht er mit der Fokussierung auf Frankreich und Australien exemplarisch auf zwei in verschiedener Hinsicht herausragende Beispiele demokratischer Online-Kontrolle eine, um aktuelle Entwicklungen der Legitimationsstrategien von Content-Regulierung zu illustrieren. Während es bei der Betrachtung von Frankreich um die Regulierung von Desinformation im Präsidentschaftswahlkampf 2017 über das *Loi infox* geht, wird im Hinblick auf Australien die Reform der Gesetzgebung diskutiert, die sich am Muster des US-amerikanischen Foreign Agents Registration Act (kurz: FARA) orientiert. Schünemann zeichnet damit einen politischen Wandel nach, der eine Reihe normativer Paradoxien zur Folge hat und damit die Informationsautonomie von Nutzer*innen potenziell erheblich schwächen können, obwohl Reformen zum Schutz der Demokratie gerade auf die institutionelle Absicherung politischer Autonomie zielen.

Über die Entstehung des Bandes: Einige Bemerkungen

Im Grunde ist die Auseinandersetzung mit der Entstehungsgeschichte eines Sammelbandes, der unterschiedliche Perspektiven verbindet, wohl nicht so interessant für die Lesenden, doch wenn die Entstehung eines Buchprojekts von einschneidenden Krisenmomenten geprägt ist, wäre es aus unserer Sicht unangemessen, diese Momente auszublenden. Der Band ist schließlich nicht nur das Resultat einer interdisziplinären Annäherung an ein aktuelles und durchaus komplexes Problemfeld, er ist auch das Ergebnis verschiedener Entscheidungen, die sich in diesem Kooperationsprozess ereignet haben. Die Entstehung des Bandes ist damit auf eine im Sommersemester 2020 an der Otto-von-Guericke-Universität Magdeburg durchgeführte interdisziplinäre Ringvorlesung zurückzuführen. Im Kreis des Organisationsteams, welches die Herausgeberschaft des Bandes abbildet, planten wir die Veranstaltung als Präsenz-

vorlesung, doch diese Planungen wurden durch die COVID-19-Pandemie im März 2020 durchkreuzt, sodass wir gemeinsam mit Referentinnen und Referenten aus ganz Deutschland, die sich bereit erklärten unserer Einladung zu folgen, in ein digitales Format und damit auf Distanzlehre gewechselt haben. Die Entscheidung zur Durchführung des Projekts in der Herausgeberschaft war maßgeblich vom Anspruch an die gemeinsame disziplinübergreifende Zusammenarbeit sowie thematische Auseinandersetzung getragen. Damit sollten neue Perspektiven für die Studierenden aber auch die Referentinnen und Referenten eröffnet und der wissenschaftliche Austausch gefestigt und fortgeführt werden. Was hätte es wohl mit den Studierenden, den Referierenden und uns gemacht, wenn eine Vorlesung zum Thema Autonomie im digitalen Zeitalter nicht auch unter diesen herausfordernden Bedingungen irgendwie realisiert worden wäre, wenn wir Veranstaltung verlegt oder abgesagt hätten? In der Ringvorlesung wurden sowohl philosophische, kultur- und bildungs- sowie erziehungswissenschaftliche Perspektiven mit jenen der Politikwissenschaft und Informatik verbunden. Die einzelnen Vorträge fokussierten unterschiedliche Ansichten dazu, wie Autonomie im digitalen Zeitalter verhandelt werden kann. Die Vorträge der Ringvorlesung wurden zudem über einen begleitenden Podcast veröffentlicht. Somit wollten wir den Studierenden sowie einem weiteren interessierten Publikum die Inhalte auf unterschiedlichen Wegen zugänglich machen und einen Beitrag zur offenen Wissenschaftskommunikation leisten. Die pandemische Lage hat sich in diesem Zeitraum verschärft und somit auch die an die Veranstaltung anschließende Publikation geprägt, weshalb der Band durch die teilweise aktualisierte Zusammensetzung durch die Beitragenden eine thematische Zuspitzung erfahren hat, die sich in der Auseinandersetzung von Fragen zur Autonomie über Momente der Selbstbestimmung und Datenpraktiken beschreiben lässt.

Die Finalisierung des Bandes erfolgt nicht nur unter den Bedingungen der Pandemie, sondern auch in einer Zeit, in der Russland die Ukraine in einem erschütternden Angriffskrieg überfällt und die Diskussion um Autonomie eine ganz neue Bedeutung erfährt. Einerseits, weil Menschen aus ihrer Heimat flüchten und andererseits weil die Masse an Informationen in den Medien ein Überangebot an Wahrheiten präsentiert und damit das Gefühl der Ohnmächtigkeit verstärkt wird. Wir möchten in diesem Band an die Menschen erinnern, denen durch Heteronomie und gewaltsame Machtausübung die Möglichkeit zu einem selbstbestimmten Leben genommen wird. Dinge die wir lange für Selbstverständlich gehalten haben, werden zur Disposition gestellt. Die Welt in der wir leben, ist nicht einfach gegeben, sie fordert uns alle dazu auf, darüber nachzudenken, wie wir sie jeden Tag neu gestalten wollen.

Literatur

Biermann, R. & Verständig, D. (Hrsg.). (2017). *Das umkämpfte Netz.* Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-15011-2.

Crawford, K., Lingel, J. & Karppi, T. (2015). Our metrics, ourselves: A hundred years of self-tracking from the weight scale to the wrist wearable device. *European Journal of Cultural Studies* 18(4–5), 479–496.

Damberger, T. (2021a). Optimal verfehlt! Dem Phänomen Selftracking bildungsphilosophisch nachgedacht. *Pädagogische Rundschau*, 75 (2), 181–188.

Damberger, T. (2021b). *Zum Verhältnis von Bildung, Nachhaltigkeit und Digitalisierung.* 30.

Dander, V. (2016). *Ideologische Aspekte von „Digitalisierung": Bd. Digitale Transformation im Diskurs* (C. Leineweber & C. de Witt, Hrsg.; S. 15). deposit_hagen. Publikationsserver der Universitätsbibliothek. https://medien-im-diskurs.de

Dander, V., Bettinger, P., Ferraro, E., Leineweber, C., Rummler, K. & Barbara Budrich. (2020). *Digitalisierung – Subjekt – Bildung: Kritische Betrachtungen der digitalen Transformation.* Budrich academic.

Duttweiler, S., Gugutzer, R., Passoth, J.-H. & Strübing, J. (Hrsg.). (2016). *Leben nach Zahlen: Self-Tracking als Optimierungsprojekt?* transcript.

Eubanks, V. (2017). *Automating inequality: How high-tech tools profile, police, and punish the poor* (First Edition). St. Martin's Press.

Foucault, M. (1997). *Die Ordnung der Dinge. Eine Archäologie der Humanwissenschaften.* Suhrkamp.

Freud, S. (2013). *Die Traumdeutung.* Nikol Verlagsgesellschaft.

Fry, H. (2019). *Hello world: Was Algorithmen können und wie sie unser Leben verändern* (S. Schmid, Übers.). C.H. Beck.

Fuchs, T. (2020). *Die Verteidigung des Menschen:* Grundfragen einer verkörperten Anthropologie. Suhrkamp taschenbuch wissenschaft.

Goldschmitt, K. E., & Seaver, N. (2019). Shaping the Stream: Techniques and Troubles of Algorithmic Recommendation. In N. Cook, M. M. Ingalls, & D. Trippett (Hrsg.), *The Cambridge Companion to Music in Digital Culture.* Cambridge University Press, S. 63–81. https://doi.org/10.1017/9781316676639.006

Hauck-Thum, U. & Noller, J. (Hrsg.). (2021). *Was ist Digitalität? Philosophische und pädagogische Perspektiven* (Bd. 1). J.B. Metzler.

Jörissen, B. & Marotzki, W. (2009). *Medienbildung. Eine Einführung: Theorie -Methoden – Analysen.* Klinkhardt.

Karcher, M. (2020). Die (kybernetische) Bändigung des Zufalls. Dataveillance und Learning Analytics als Herausforderung erziehungswissenschaftlicher Reflexion: Versuch einer Technikfolgenabschätzung. In D. Fickermann, V. Manitius, & M. Karcher (Hrsg.), *„Neue Steuerung" – Renaissance der Kybernetik?* Waxmann Verlag GmbH, S. 151–167. https://doi.org/10.31244/9783830991618.10

Knuth, D. E. (1968). *The art of computer programming.* Addison-Wesley.

Leineweber, Christian, M. (2021). Digitale Aufklärung? Datenkritik und Urteilsfähigkeit. In C. Leineweber & C. de Witt (Hrsg.), *Algorithmisierung und Autonomie im Diskurs – Perspektiven und Reflexionen auf die Logiken automatisierter Maschinen.* Hagen University Press, S.25–154. https://medien-im-diskurs.de

Marx, K. (2005). *Ökonomisch-Philosophische Manuskripte*, hrsg. von Barbara Zehnpfennig. Meiner Verlag.

Nguyen, C. T. (2020). ECHO CHAMBERS AND EPISTEMIC BUBBLES. Episteme, 17(2), 141–161. https://doi.org/10.1017/epi.2018.32

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information.* Harvard University Press.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you.* Penguin Press.

Quante, M. (2011). Autonomie. In S. Jordan & C. Nimtz (Hrsg.), *Lexikon Philosophie. 100 Grundbegriffe.* Reclam, S. 47–49.

Reckwitz, A. (2003). Grundelemente einer Theorie sozialer Praktiken / Basic Elements of a Theory of Social Practices: Eine sozialtheoretische Perspektive / A Perspective in Social Theory. *Zeitschrift Für Soziologie*, *32*(4), 282–301. https://doi.org/10.1515/zfsoz-2003-0401

Reinisch, M. (2020). Big Data und Algorithmen: Instrumente einer neuen kybernetischen Steuerung an Schulen? In D. Fickermann, V. Manitius, & M. Karcher (Hrsg.), *„Neue Steuerung" – Renaissance der Kybernetik?.* Waxmann Verlag GmbH, S. 134–150. https://doi.org/10.31244/9783830991618.09

Rosenblat, A. (2018). *Uberland: How algorithms are rewriting the rules of work.* University of California Press.

Roth, G. (1996). *Das Gehirn und seine Wirklichkeit:* Kognitive Neurobiologie und ihre philosophischen Konsequenzen. Suhrkamp taschenbuch wissenschaft.

Roy, S., Menapace, W., Oei, S., Luijten, B., Fini, E., Saltori, C., Huijben, I., Chennakeshava, N., Mento, F., Sentelli, A., Peschiera, E., Trevisan, R., Maschietto, G., Torri, E., Inchingolo, R., Smargiassi, A., Soldati, G., Rota, P., Passerini, A., Demi, L. (2020). Deep Learning for Classification and Localization of COVID-19 Markers in Point-of-Care Lung Ultrasound. *IEEE Transactions on Medical Imaging*, *39*(8), 2676–2687. https://doi.org/10.1109/TMI.2020.2994459

Rummler, K., Döbeli Honegger, B., Moser, H., & Niesyto, H. (2016). Editorial: Medienbildung und informatische Bildung – quo vadis? *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, *25*, 1–6. https://doi.org/10.21240/mpaed/25/2016.10.24.X

Schiefner-Rohs, M., Hofhues, S., & Breiter, A. (2021). Editorial: Datengetriebene Schule. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, *44*, i–xii. https://doi.org/10.21240/mpaed/44/2021.10.25.X

Stalder, F. (2016). *Kultur der Digitalität* (Erste Auflage, Originalausgabe). Suhrkamp.

Waldmann, M., & Wunder, M. (2021). Es empfiehlt sich ‚von selbst'—Bildungssoziologische Überlegungen zur Transformation von Autonomieverhältnissen durch Recommender-Systeme in der Hochschullehre. In C. Leineweber & C. de Witt (Hrsg.), *Algorithmisierung und Autonomie im Diskurs – Perspektiven und Reflexionen auf die Logiken automatisierter Maschinen.* Hagen University Press, S. 68–101. https://medien-im-diskurs.de

Zweig, K. A. (2019). *Ein Algorithmus hat kein Taktgefühl: Wo künstliche Intelligenz sich irrt, warum uns das betrifft und was wir dagegen tun können.* Heyne.

Berechnete Unbestimmtheit: Paradoxien der Freiheit im digitalen Zeitalter

Dan Verständig, Janne Stricker

1 Einleitung

Der Wunsch nach Freiheit ist untrennbar mit dem Wesen des Menschen verbunden. Doch was genau Freiheit wirklich ausmacht, ist meist nur schwer zu fassen und hochgradig ambivalent. Der folgende Beitrag macht es sich zur Aufgabe, Konturen von Freiheit im digitalen Zeitalter aus einer Perspektive der Medienbildung in interdisziplinär anschlussfähiger Weise nachzuzeichnen. Dabei werden anhand des hier konturierten Freiheitsbegriffs Paradoxien herausgearbeitet, die sich auf das Zusammenspiel von Mensch und Maschine beziehen und gesellschaftliche Konsequenzen beleuchten. Wir sprechen dabei von der berechneten Unbestimmtheit. Dies ist unser Vorschlag einer Beschreibung für die paradoxalen Strukturen, die sich aus Freiheit und Digitalität ergeben[1]. Begründen lässt sich diese Perspektivsetzung damit, dass grundlegende Fragen danach, ob und wie man die richtige Entscheidung trifft, heute von einer neuen Unübersichtlichkeit (Stalder 2016) geprägt sind. Aus dieser Unübersichtlichkeit entsteht zugleich eine „große Unordnung“ (S. 114), die Menschen dazu auffordert und herausfordert, „eigene Ordnungen erstellen zu können, als auch die Verpflichtung, eigene Ordnungen erstellen zu müssen, ob er auf diese Aufgabe nun vorbereitet ist oder nicht“ (Stalder 2016, S. 116). Damit steht die Notwendigkeit zur Herstellung von Ordnungen im Zusam-

1 Khurana und Menke (2019) besprechen aus philosophischer Perspektive Paradoxien der Autonomie und bringen in ihrem Sammelband verschiedene Autor*innen zusammen, die auch den Freiheitsbegriff in den Blick nehmen. Baer und Sacksofsky (2018) verhandeln aus einer gendertheoretische Position die Fragen der Gleichstellung und Freiheit. Nida-Rümelin (2005) befasst sich grundlegend mit der menschlichen Freiheit und beantwortet dabei die Fragen danach, warum es keine Verantwortung ohne Freiheit gibt und warum Menschenwürde auf Freiheit beruht.

menhang mit der individuellen Herstellung von Orientierung. Diese Herausforderungen werden in diesem Beitrag im Kontext von Freiheit, algorithmischen Berechnungen und Daten in den Mittelpunkt gestellt. Dabei leitet uns die Frage, wo die Grenzen und Reichweiten des Freiheitsbegriffs im digitalen Zeitalter liegen. Zur Beantwortung dieser Fragen werden wir eine bildungstheoretische Perspektive einnehmen, indem wir Freiheit und die Gesetzmäßigkeiten des Digitalen ins Verhältnis miteinander setzen und die daraus resultierenden paradoxalen Strukturen diskutieren.

Die Bestimmung des Verhältnisses von Freiheit und Bildung hat in der Erziehungswissenschaft eine lange Tradition, die sich grundlegend mit der Entfaltung des Menschen und seiner Selbstbestimmung befasst (Vieweg & Winkler 2012). Trotz dieser langen Tradition – oder gerade deswegen – lassen sich die Grenzen und Reichweiten des Freiheitsbegriffs angesichts der technologischen Entwicklungen und gesellschaftlichen Veränderungsdynamiken erneut bestimmen. Um analytisch zu erfassen, wie Freiheit im digitalen Zeitalter konturiert werden kann, wollen wir eine erste Bestimmung des Freiheitsbegriffs unabhängig von der technologischen Bedingung skizzieren und in Analogie zu Höffe (2018) Kernmerkmale des Diskurses herausstellen. In einem zweiten Schritt werden diese Merkmale dann auf die digitale Welt und somit die digitalen Technologien und ihre Infrastrukturen bezogen. Hier liegt das Hauptaugenmerk auf Algorithmen und digitalen Daten, die mal mehr, mal weniger bewusst erstellt, geteilt und damit auch kontextualisierbar und verstehbar gemacht werden. Dabei ist unsere Perspektive von der Annahme geleitet, dass Daten performativ sind, das heißt, sie sind nicht nur ein Abbild der Welt, sie machen Welt. Schließlich werden daraufhin die paradoxalen Strukturen von Freiheit in der digitalen Welt diskutiert, wobei gefragt wird, welche Türen von Freiheit und Autonomie sich durch die Digitalität öffnen bzw. schließen. Ziel des Beitrags ist es einerseits aufzuzeigen, inwiefern digitale Technologien und ihre Infrastrukturen schließlich auf die individuelle Freiheit einwirken.

2 Welche Freiheit?

Wenn man von Freiheit spricht, dann ist zunächst nicht sofort klar, welche Freiheit denn eigentlich gemeint ist. In unserer strukturtheoretischen Betrachtung wird Freiheit als Prozess verstanden, der durch ständige Suchbewegungen immer wieder neu ausgehandelt wird. Um dieses Verständnis deutlich und an unsere Überlegungen anschlussfähig zu machen, wird im Folgenden mit Höffe (2018, S. 12) ein Freiheitsbegriff in fünf Themenbereichen konturiert, der strukturtheoretische Anschlüsse auch mit Blick auf die Digitalität zulässt. Die

Bereiche erstrecken sich (1) von den Naturzwängen über (2) Wirtschaft und Gesellschaft, über (3) Wissenschaft und Kunst zur (4) Welt des Politischen und schließlich (5) der personalen Freiheit. Die Freiheit von Naturzwängen zeichnet sich unter anderem über das menschliche Bestreben aus, sich über die Gegebenheiten der Natur hinwegsetzen und selbst schöpferisch tätig zu werden. Freiheit richtet sich dann einerseits auf die Bändigung der zerstörerischen Naturgewalten, die Domestizierung von Tieren oder die Bekämpfung von Krankheiten. Diese Bestrebungen sind stets verbunden und eng verwoben mit technologischen Entwicklungen und damit ein Signum für die jeweiligen Epochen. Typisch für die Moderne oder die Übergangsphase in die Spätmoderne ist hier die enorme Ausweitung, „Steigerung und Verdichtung von Gefahrenbekämpfung, Arbeitserleichterung und Ertragssteigerung" (Höffe 2018, S. 14 f.), jedoch auch die damit einhergehende Naturbelastung. Der Mensch macht sich die Natur zu eigen und nutzt natürliche Ressourcen, um daraus technologische Werkzeuge und Entwicklungen abzuleiten und herzustellen, die dann wieder an die Natur und ihre Ressourcen zurückgebunden sind.

Nimmt man das Medienhandeln zum Anlass, wird diese zyklische Abhängigkeit deutlich, während in der nördlichen Hemisphäre über den technologischen Fortschritt gesprochen und gestritten wird, gräbt sich das Digitale sprichwörtlich in die Erdschichten der südlichen Hemisphäre ein und zeichnet damit ganz gravierende Einschnitte der praktischen Produktionsbedingungen, denn die für die Technologieherstellung nötigen Mineralien wie beispielsweise Cobalt, Coltan oder Lithium (Parikka 2015) lassen soziale Schieflagen und Ausbeutungsstrukturen deutlich werden, die von Gesundheitsgefährdung über Kinderarbeit reichen und ganz grundlegend auch den Umweltschutz betreffen (Grünberger 2021, S. 222).

Das Überwinden von naturgesetzlichen Grenzen und das zugrunde liegende Mantra der Grenzüberwindung bildet sich nicht nur im Umgang mit der Technologie ab, es formt auch den Menschen, der sich stets im Verhältnis zur Natur sieht und darüber neu definiert, seien es genetische Manipulationen oder andere biologische Eingriffe und Anpassungen (Damberger 2015, O'Connell 2017, Schenk & Karcher 2018). Diese Eingriffe und Entwicklungen sind geprägt von ethischen Diskursen. Es geht dabei in erster Linie um das Spannungsverhältnis von individueller Freiheit einerseits und den technologischen Entwicklungen sowie gesetzlichen Rahmenbedingungen andererseits. Damit wird eine weitere Facette des Freiheitsbegriffs relevant, die Höffe (2018) in seinen Schwerpunktsetzungen ebenfalls berücksichtigt: Es handelt sich um eine Differenzierungsleistung von Freiheit in Wirtschaft und Gesellschaft, denn der Mensch ist kein „isoliertes Individuum", das in „wirtschaftlicher, gesellschaftlicher, kultureller und politischer Hinsicht bindungslos existiert", sondern ganz im Gegenteil auf vielfältige Bindungen angewiesen, die keinen Selbstzweck haben, sondern letztlich dem Menschen zugutekommen sollen und dessen freie Zustimmung verdienen (ebd., S. 15). Damit ergibt sich ein

Spannungsfeld, was sich einerseits aus einer neoliberalen Marktlogik an der Herausbildung des Individuums festmachen lässt und andererseits soziale Fragen der Gerechtigkeit nicht nur regional, sondern global ins Verhältnis setzt, jedoch im Hinblick auf den Sozialstaat an normative sowie regulative Kriterien gebunden ist. Die Freiheit des Marktes steht damit der sozialstaatlichen Ausrichtung einer demokratisch verfassten Gesellschaftsordnung gegenüber. Für die individuelle Freiheit werden damit wichtige Bezugspunkte deutlich, die es sich im Hinblick auf digitale Technologien und Daten zu diskutieren lohnt.

Die individuelle Freiheit ist jedoch nicht nur über die Willensfreiheit definiert, sondern auch in Vereinigung mit Naturmechansimen zu sehen. In der reinsten Form der Menschheit wirkt die Natur mit ihren Gegebenheiten und Gesetzmäßigkeiten auf den Menschen und dessen Leben ein. Autonomie kann daher nicht ohne Weiteres allein auf den Moment der vernünftigen und sittlichen Selbstbestimmung reduziert werden, da die Menschen im Alltag einerseits auf andere Menschen und andererseits auf die Natur angewiesen sind (vgl. Damberger 2019, S. 328). Dennoch wird für die Konturierung von Freiheit gerade dieses Verhältnis zur Autonomie dann relevant, wenn es um das Verhältnis von individueller Willensbildung und Handlungsfreiheit geht und damit die Umwelt einbezogen wird. Dann, wenn Menschen etwas von sich selbst anstoßen, dann, wenn sich Menschen ihrer selbst bewusst werden, wird individuelle Freiheit nicht als Zustand, sondern als Prozess von paradoxalen Strukturen und der Herausforderung zur Vereinigung normativer Gesetzmäßigkeiten und gesellschaftlicher Aushandlungen geprägt.

Blickt man nochmals auf die Epoche der Aufklärung, dann wird Freiheit auch im Hinblick auf Wissenschaft und Kunst in den Mittelpunkt gestellt. Der freie Zugang zu Informationen und die Bedeutung von Wissen und Wissensarbeit sind auch heute zentrale Kategorien, wenn es um Freiheit und Erkenntnis geht, doch sie sind trotz ihrer Verankerung im Grundgesetz nach wie vor umkämpfte Felder. Neben der Wissenschaft kann Kunst, sofern sie sich nicht den Zwängen der Kommerzialisierung unterwirft, zum Ursprung von Kreativität und Originalität und damit ein hochdynamischer Raum für Freiheit und die ständige Aktualisierung und Befragung bestehender gesellschaftlicher Vorstellungen und Werte gelten. Damit wird auch die Politik der Freiheit angesprochen, wie Höffe (2018) als den vierten Schwerpunktbereich herausgestellt und am Beispiel der Globalisierung im Zusammenhang zum Demokratieprinzip verdeutlicht. Die politische Freiheit umfasst die Legitimierung der freiheitlichen Demokratie durch die Gewährung von Zugang zu gesellschaftlichen und politischen Diskursen. Damit werden die Rechte von Bürgerinnen und Bürgern zur Partizipation an der Gesellschaft beispielsweise durch das Wahlrecht sichergestellt, indem die Menschen ihren Willen in demokratischer Weise in den politischen Willensbildungsprozess einbringen können. Politische Freiheit meint auch die Sicherstellung und Durchsetzung weiterer Grundrechte, wie die Meinungsfreiheit und Versammlungsfreiheit. Damit steht Freiheit im

ganz natürlichen Spannungsverhältnis von gesellschaftlichen Konventionen und Normen sowie individuellen Werten. Das Bedürfnis nach individueller Freiheit muss im rechtsstaatlichen Sinne im Verhältnis zur gesellschaftlichen Ordnung gesehen werden. Freiheit der Einzelnen kann nicht zu Lasten der Anderen sichergestellt werden. Dieses Spannungsverhältnis der personalen beziehungsweise individuellen Freiheit wird besonders dann deutlich, wenn man Freiheit entlang von Gerechtigkeit und Grenzen verortet. Der Mensch ist im Sinne der Aufklärung durch seine Sprach- und Vernunftbegabung zur Freiheit fähig, jedoch ist dies zumindest bei Kant immer auch die Notwendigkeit, sich der Kraft seiner Vernunft bewusst zu werden. Damit wird die Differenz von Freiheit und Pflicht nivelliert. Mit der Konturierung der hier kursorisch rekonstruierten Themenbereiche nach Höffe (2018) haben wir einerseits ein Konzept von Freiheit skizziert, dass auch heute noch Einflüsse der Aufklärung erkennen lässt. Andererseits wird eine Prozesshaftigkeit deutlich, die sich in zweierlei Hinsicht abzeichnet: *Erstens* wird Freiheit in historischer Entwicklung immer wieder in Abhängigkeit zu gesellschaftlichen, politischen und kulturellen Rahmenbedingungen thematisiert. *Zweitens* ist Freiheit und besonders mit Blick auf die individuelle Handlungsfähigkeit, Willensbildung- und Entscheidungsfindung selbst ein Prozess, der den Menschen in seiner grundlegenden Verfassung von inneren und äußeren Einflüssen prägt und zumindest die individuelle Freiheit damit immer auch im Spannungsfeld zu normativen Ordnungen steht. Das hier konturierte Konzept von Freiheit richtet sich nicht allein auf die individuelle Freiheit, sondern stellt insofern Bezugspunkte zu Feldern der Wissenschaft, Politik und Gesellschaft her, als dass hierdurch Anknüpfungen zu einer zeitgemäßen Bestimmung von Freiheit unter der digitalen Bedingung vorgenommen werden können.

Wenn wir also von Freiheit sprechen, dann rücken wir in erster Linie die Entscheidungs- und Handlungsfreiheit in den Mittelpunkt. Dabei ergibt sich für den Fortgang des Beitrags die Grundproblematik, dass im digitalen Zeitalter zunächst danach gefragt werden muss, ob oder inwiefern Entscheidungen selbst- oder fremdbestimmt erfolgen. Der Mensch ist aufgefordert, sich den algorithmischen Strukturen gegenüber zu verhalten, ohne sich ihnen jedoch gegenüberstellen zu können, da sie aufgrund ihrer Architekturen weitestgehend unverfügbar bleiben.

3 Freiheit im digitalen Zeitalter

Will man die zuvor entfaltete Konzeption von Freiheit entlang der zeitgenössischen technologischen Entwicklungen rund um das Internet und die Digitalisierung bestimmen, ist zunächst eine Einordnung dessen nötig, wie sich der

digitale Wandel darstellt. Dies erfolgt in drei Schritten. In einem ersten Schritt soll eine grundlegende Differenz zwischen Digitalisierung und Digitalität herausgearbeitet werden, bevor die Regeln und die Gesetzmäßigkeiten des Digitalen vertieft und schließlich die individuelle Herstellung von Orientierung im Spannungsfeld von Unbestimmtheit und Berechnung strukturtheoretisch betrachtet wird.

3.1 Digitalisierung und Digitalität

Mit dem Begriff der Digitalisierung werden seit einigen Jahren tiefgreifende Transformationsdynamiken deutlich, die sich nicht nur über einzelne Felder erstrecken, sondern im gesamtgesellschaftlichen Kontext stehen. Für die Arbeitswelt ergeben sich durch die technologischen Entwicklungen der Digitalisierung neue Anforderungen an das, was gelernt werden muss, welcher Logik das Lernen folgt und was man Arbeit nennt. Auf die Lebenswelt wirkt sich die Digitalisierung insofern aus, als dass aus den vielfältigen Kanälen zur Informationsbeschaffung und den digitalen medialen Formaten ein neuer Möglichkeitsraum entsteht, der potenzielle Teilhabe an gesellschaftlichen Prozessen in einer ganz neuen Weise bietet. Ob und inwieweit die Potenziale zur Teilhabe der Digitalisierung tatsächlich entfaltet werden, ist umstritten, denn die Technologien sind eben nicht neutral, sie ermöglichen keinen uneingeschränkten Zugriff auf Informationen, vielmehr sind sie grundlegend ideologieanfällig, wie Eubanks (2018), Zuboff (2019) oder Daub (2020) in kritischen Studien zu Ungleichheit, Überwachung und Kapitalismuskritik herausgearbeitet haben. Damit ist unstrittig, dass technologische Entwicklungen der Digitalisierung tief in die Lebenswelt der Menschen reichen. Beispielsweise wenn Algorithmen dafür eingesetzt werden, um Vorhersagen zu treffen, die bereits auf Basis ihrer Konzeption zur Verschärfung sozialer Schieflagen führen. Dabei handelt es sich um einen sehr kleinen Kreis von Menschen, die in der Position und Verantwortung sind, die Plattformen durch das Design, die Entwicklung und die Wartung mitzugestalten. Dieser Kreis von Menschen ist heute weitestgehend im und um das Silicon Valley konzentriert, da dort die großen Tech-Firmen beheimatet sind. Es handelt sich gewissermaßen um das Epizentrum der Tech-Branche und dementsprechend ist auch das Denken über Technologie und technologische Entwicklungen von den Menschen geprägt, die ihre Ideen und Visionen in die Realität und in Form von Plattformen oder anderen Angeboten umsetzen, wie Daub (2020) entlang einer kritischen Analyse über die Rhetorik der Tech-Branche unter Berücksichtigung historischer Linien herausgearbeitet hat. Das Denken über technologische Entwicklungen und Fortschritt gibt Aufschluss darüber, wie schließlich Prozesse und Ideen in Hard- und Software umgesetzt werden. Damit ist die ideengeschichtliche Annäherung an einen Standort, der vermeintlich keine Geschichte hat, nicht nur dann relevant, wenn

es um Fortschritt und Disruption geht, sondern auch ein Anknüpfungspunkt für die Konturierung von Freiheit. Während Plattformen wie Lyft oder Uber (Rosenblat 2018) damit werben, dass sie die Taxifahrt revolutioniert haben, hat sich am Erlebnis für die Kundinnen und Kunden in der Realität nicht viel geändert, die Unternehmen haben jedoch grundlegende Kategorien von Arbeitsverhältnissen verschoben, indem prekäre Konstellationen und Scheinselbstständigkeiten zur Basis dieser Ideen wurden (vgl. Daub 2020, S. 106). Plattformen der Sozialen Medien sind keine neutralen Instanzen, die versehentlich zu sozialen und politischen Polarisierungen und Konflikten führen, sie begünstigen diese Entwicklungen, indem sie Dateninfrastrukturen in einer bestimmten Weise bereitstellen und diese über ihre Geschäftsmodelle ausrollen, die sich dann in die privatesten Sphären der Lebenswelt einschreiben (Zuboff 2019). Was sich hier beobachten lässt, ist eine Entkopplung von technologischen Entwicklungen einerseits und dem Denken sowie der Rhetorik über diese Zusammenhänge andererseits. Für die Konturierung von Freiheit im digitalen Zeitalter sind diese Entwicklungen deshalb wichtig, da diese Entkopplung hier zu einem wichtigen Angelpunkt für die Herstellung der individuellen Freiheit wird, der sich über unterschiedliche mediale und datenbasierte Praktiken abbilden lässt. Dennoch wird durch eine Metaperspektive auf Digitalisierung wenig über die tatsächlichen Auswirkungen digitaler Technologien auf arbeits- und lebensweltliche Zusammenhänge gesagt.

Mit der „Kultur der Digitalität" problematisiert Stalder (2016) die Digitalisierung und die Auswirkungen auf soziales und politisches Handeln aus einer kulturwissenschaftlichen Perspektive. Seine Überlegungen basieren auf dem Umstand, dass immer mehr Menschen die Möglichkeit haben, sich durch digitale Technologien zu beteiligen und dies auch tun. Den Begriff der Digitalität begründet Stalder (2016) mit der These, dass das Internet inzwischen „für viele eine Infrastruktur [ist], die sie wie Wasser- oder Stromnetze als gegeben voraussetzen" (Stalder 2016, S. 92). Damit wird dem Umstand Rechnung getragen, dass Digitalisierungsprozesse so weit vorangeschritten sind, dass sie der dominante Kulturraum sind. Digitalität sei damit so dominant und allgegenwärtig (vgl. Stalder 2016, S. 94), dass sich immer mehr Menschen an kulturellen Prozessen beteiligen und somit immer weitere Dimensionen der Existenz zu Feldern kultureller Auseinandersetzungen werden. Dementsprechend werde soziales Handeln in zunehmend komplexere Technologien eingebettet, „ohne die diese Prozesse kaum zu denken und schon gar nicht zu bewerkstelligen wären" (Stalder 2016, S. 11). Das Internet wird damit zum natürlichen Bestandteil lebensweltlicher Zusammenhänge und Aushandlungsprozesse sowie die Grundlage für viele Menschen, sich über ganz unterschiedliche Kanäle und Angebote Informationen zu beschaffen. Die heutigen Möglichkeiten der Informationsbeschaffung und des Austauschs sind nicht das Ergebnis einer einzelnen technologischen Entwicklung, sondern vielmehr sukzessiver Entwicklungsphasen, die ineinandergreifen und sich über unterschiedliche Bereiche

erstrecken, ineinanderwirken und damit ein Wechselverhältnis von technologischer Entwicklung und sozialen, politischen und kulturellen Dynamiken beschreiben.

Nach Stalder (2021) könne man von Digitalität ungefähr seit dem Jahr 2000 sprechen (vgl. ebd., S. 4). Damit wird eine Entwicklung hin zu einem Zustand beschrieben, der sowohl das Materielle als auch das Immaterielle, sowohl das Analoge als auch das Digitale beinhaltet und keine kategorische Trennung dieser Bereiche vornimmt, denn unter den Bedingungen der Digitalität verschwindet das Analoge nicht, es bekommt vielmehr eine neue Bedeutung zugeschrieben. Besonders deutlich wird das durch Smartphones, die inzwischen zum alltäglichen Begleiter entwickelt haben. Die es uns erlauben, zu jeder Zeit an jedem Ort Fotos und Videos aufzunehmen, Notizen zu verfassen und mit anderen Menschen spontane Eindrücke, Erlebnisse, aber auch Krisenmomente zu teilen. Auch wurden mit der Intensivierung immaterieller Arbeitsleistungen – sofern man diese Trennung tatsächlich aufmachen kann – die Fabrikjobs nicht einfach abgelöst, sie wurden ausgelagert und damit zu einem Bestandteil einer globalen Produktions- und Vermarktungskette. Digitalität bezeichnet damit jenes „Set von Relationen, das heute auf Basis der Infrastruktur digitaler Netzwerke in Produktion, Nutzung und Transformation materieller und immaterieller Güter sowie in der Konstitution und Koordination persönlichen und kollektiven Handelns realisiert wird" (Stalder 2016, S. 14). Dieses Set von Relationen wird von Stalder (2016) in drei grundlegende Eigenschaften unterteilt. Es handelt sich um Referentialität, Gemeinschaftlichkeit und Algorithmizität. Referentialität meint „[…] die Nutzung bestehenden kulturellen Materials für die eigene Produktion […] als zentrale Eigenschaft vieler Verfahren, mit denen sich Menschen in kulturelle Prozesse einschreiben" (ebd., S. 13). Dazu gehöre ebenfalls ein bewusstes in Referenz Treten mit digitalen Materialien, was geprägt ist von basalen Akten der Bedeutungsproduktion und Selbstkonstitution. Gemeinschaftlichkeit als zweite zentrale Eigenschaft meint, dass Kultur „verstanden als geteilte soziale Bedeutung" (ebd., S. 95), als etwas Gemeinsames entsteht. Durch Interaktionen und soziale Aushandlungen gemeinsame Bedeutungen zugeschrieben werden. Es geht darum, dass durch die Gemeinschaftlichkeit kollektive Rahmen der Interpretation geschaffen werden. Gleichzeitig habe dies aber auch zur Folge, dass die vorhandenen Ressourcen und kulturellen Güter für alle zugänglich und erreichbar seien. Algorithmizität schließlich verweist auf die automatisierten Entscheidungsverfahren,

> „die den Informationsfluss reduzieren und formen, so dass sich aus den von Maschinen produzierten Datenmengen Informationen gewinnen lassen, die der menschlichen Wahrnehmung zugänglich sind zu Grundlagen des singulären und gemeinschaftlichen Handelns werden können" (ebd., S. 13).

Ohne die Algorithmen wären wir nach Stalder blind (vgl. ebd.). Algorithmen sind damit für die Herstellung von individueller Orientierung bedeutsam, denn

sie stellen einen Rahmen dar, der der Gemeinschaftlichkeit noch vorgelagert ist, gewissermaßen die Bedingung der Möglichkeit von Bedeutungskonstitution (vgl. Noller 2021, S. 42). Stalder (2016) fasst dies in einem dialektischen Verhältnis, denn algorithmische Selektionsmechanismen, wie beispielsweise die Google-Suche, schaffen Freiheiten, indem sie die „Voraussetzungen der persönlichen Handlungsfähigkeit generieren" und sie schaffen damit aber auch neue Abhängigkeiten, indem sie die „informationelle) Welt vorsortieren und zugänglich machen" (Stalder 2016, S. 96). Für die Konturierung von Freiheit ist es daher auch wichtig festzuhalten, dass Algorithmen auf das Handeln der Menschen einen grundlegenden Einfluss haben, dem man sich vor dem Hintergrund der Algorithmizität kaum entziehen kann, denn Algorithmen prägen den Alltag der Menschen in unterschiedlichen Kontexten, ganz gleich, ob es den Menschen bewusst ist oder nicht. Menschen nutzen Algorithmen beispielsweise dann bewusst, wenn sie gezielt nach Informationen suchen oder Daten beispielsweise durch Praktiken der Selbstvermessung produzieren und diese selbst prüfen. Ein mehr oder weniger bewusster Umgang mit algorithmischen Filtermechanismen ist beispielsweise die Nutzung von Hashtags in den Sozialen Medien, um sich an thematischen Debatten zu beteiligen und seine Position mitzuteilen. Demgegenüber stehen weitestgehend unbewusste und zunächst unsichtbare Wirkweisen von Algorithmen und Daten, beispielsweise wenn die Nutzung von Diensten durch Dritte ausgewertet wird. Unternehmen werten das Nutzungsverhalten der Menschen nicht selten aus, um eigene Interessen und die Monetarisierung der Dienste voranzutreiben. Manchmal sind die Interessen der Unternehmen im Einklang mit den Nutzungsweisen der Menschen – manchmal nicht. Häufig ist bislang noch unklar, in welchem Umfang Daten von den Unternehmen erhoben, ausgewertet und wie diese Daten weiterverarbeitet werden. Unklaur bleibt damit auch, inwiefern die Umwelt der Menschen dadurch strukturiert wird. Das Spektrum reicht von der Anzeige individueller Werbung, über die Empfehlung von Freundesfreunden hin zur individuellen Strukturierung der persönlichen Timeline. Die persönliche Freiheit wird damit zu einer Freiheit der individuellen Persönlichkeitsentfaltung, die geprägt ist von den Möglichkeiten des eigenen Ausdrucks in den Sozialen Medien. Reckwitz (2017) beschreibt in seiner Theorie der „Gesellschaft der Singularitäten" die Inszenierung von Einzigartigkeit und die Herstellung von öffentlicher Sichtbarkeit, die Suche nach Anerkennung durch Einzigartigkeit, wie sie bei Influencer*innen thematisiert werden kann. Habermas (2021) grenzt dieses Singularitätsversprechen jedoch von Individualisierung insofern ab, als dass der Individualisierungsbegriff auf die „lebensgeschichtlich erworbene Unverwechselbarkeit" abzielt (ebd., S. 494). Vor dem Hintergrund der Digitalität bleibt dann zu diskutieren, inwiefern Singularisierung und Individualisierung doch Überschneidungen aufweisen, wenn die Regeln des Digitalen auch die Lebenswelten der Menschen beeinflussen.

3.2 Regeln und Gesetzmäßigkeiten des Digitalen

Der Rechtswissenschaftler Lawrence Lessig gibt mit seiner recht provokanten These „Code is Law“ (1999) schon recht früh einige wichtige Hinweise dahingehend, inwiefern technologische Entwicklungen der digitalen Vernetzung als Dreh- und Angelpunkt für die Herausbildung neuer Gesetzmäßigkeiten beschrieben werden können. Dabei legt er einen weiten Codebegriff zugrunde, der sowohl Hard- als auch Software, damit auch die technologischen Infrastrukturen und Architekturen einbezieht. Lessig (1999) entkräftet dabei den Mythos, dass der Cyberspace neue Freiheiten mit sich bringe. Es ein Irrtum, daran zu glauben, dass das Internet mit all seinen unterschiedlichen Diensten und Protokollen ein freier Raum sei, denn der Code sei nicht festgelegt, sondern werde wie Gesetze ausgelegt, woraus sich erst die Gesetzmäßigkeiten ergeben. Die Auslegung dieser Gesetzmäßigkeiten erfolge jedoch nicht in erster Linie durch die Nutzenden, sondern durch die Entwickler:innen und diejenigen, die Kontrolle über die Infrastrukturen haben. Dabei stellt er sich gegen die Idee, dass das Internet als Kulturraum ein natürlich gewachsener Raum sei. Digital vernetzte Architekturen hätten nichts mit Natur an sich zu tun (vgl. ebd., S. 31), sie seien von Menschen gedacht und gemacht. Der Mensch versucht, sich über Technik selbst neu zu erfinden, sich zu verbessern, die natürlichen und selbst auferlegten Grenzen sowie Gesetzmäßigkeiten zu überwinden, um seinem Streben nach Freiheit nachzukommen. Damit verändert er jedoch das Verhältnis zur Natur und Umwelt, in der lebt, was ihn auf sich selbst und die Naturzwänge zurückwirft und gleichzeitig neue Zwänge durch Technik hervorbringt.

Hier lässt sich an die von Stalder (2016) entwickelte Linie zur Digitalität anschließen, denn die digitalen Architekturen haben um diese Zeit an Popularität gewonnen und ihre Verbreitung durch steigende Vernetzung erfahren. Manch vernetzte Architekturen werden dabei, so Lessig (1999) mehr reguliert und manche weniger. Bei Stalder (2016) ist Algorithmizität ein grundlegender Ankerpunkt für Digitalität und damit ein strukturiertes und strukturierendes Moment. Dabei bezieht er sich unter anderem auf Algorithmen, die zur Selektion und Suche von Informationen genutzt werden und jene, die das soziale Miteinander in vermittelten Kontexten strukturieren. Das Spektrum der algorithmischen Systeme, die in Form der digitalen Architekturen in den Alltag der Menschen eingeschrieben sind, übersteigt Suchalgorithmen und erstreckt sich indirekt auch auf viele andere Kontexte, wie die Nutzung von öffentlicher Infrastruktur, die von algorithmischen Strukturen geprägt ist oder den Einkauf in einem Supermarkt, mit vernetztem Kassensystem. Algorithmen zeichnen sich trotz ihrer Heterogenität stets als bestimmte Handlungsanweisungen in einer Programmiersprache ab, die dann in Maschinensprache übersetzt und durch Computer in irgendeiner spezifischen Form ausgeführt werden. Dabei setzen sich die Handlungsanweisungen lediglich aus den vorstrukturierten und

implementierten Bedingungen zusammen. Wie die Ausführung tatsächlich erfolgt, hängt in erster Linie von den Daten ab, die dem Algorithmus zugrunde liegen. Die Algorithmen, die hier im Mittelpunkt der Betrachtung stehen, sind dementsprechend mehr und mehr datengetrieben und ihre *Lösung* oder *Ergebnisse* der Berechnung dabei weniger das Resultat der *Programmierenden* und ihrer implementierten Ideen (vgl. Matzner 2019, S. 3). Algorithmische Systeme verarbeiten und reproduzieren gesellschaftlich ausgehandelte Werte mehr und mehr durch Daten und manchmal wird das auch sichtbar. Hieraus ergeben sich zwei Konsequenzen für die Betrachtung von Algorithmen und Daten im Hinblick auf Freiheit: Wenn nicht klar ist, wie algorithmische Systeme tatsächlich wirken, fällt es erstens schwer Erklärungen darüber anzustellen, inwieweit sie als Black Boxes auf menschliche Entscheidungs- und Handlungsprozesse Einfluss haben. Zweitens genügt es nicht bloß, die Black Boxes zu öffnen, um Verzerrungen und Reichweiten von Algorithmen und Daten zu bestimmen, vielmehr müsse dieser Prozess der Transparenzschaffung, so Matzner (2017, S. 45), in einen größeren sozialen, institutionellen und technischen Rahmen eingebettet werden, welcher Machtverschiebungen und Subjektivitäten analysiert. Folglich reicht weder die Perspektive auf Algorithmen als irgendwie unabhängigen, neutralen Akteur gegenüber dem Menschen aus, noch genügt es, Softwarecode als neutrales Produkt zu betrachten, was in seiner Ausführung strikt genau das macht, wofür es gedacht wurde. Algorithmen sind vielmehr eingebettet in soziale und kulturelle Praktiken und damit ebenso wie Daten performativ. Sie entfalten ihre Komplexität im Zusammenhang mit ihren Implementationen, die Software sowie ihre unterschiedlichen und vielfältigen Anwendungskontexte. Algorithmen und Daten sind dadurch eben nicht objektiv und fehlerfrei, wie ihnen meist unterstellt wird. Dementsprechend haben sie auch direkte Auswirkungen auf die Herstellung von individuellen, aber auch kollektiven Orientierungsrahmen.

3.3 Herstellung von Orientierung und der Umgang mit Unbestimmtheit

Die Herstellung von Orientierung ist eine Notwendigkeit des Menschen, sich in einer Gesellschaft zurechtzufinden und handeln zu können. Dabei ist Orientierung nicht gleichzusetzen mit selbstbestimmtem Handeln, erst der Modus der Selbstreflektion befähigt den Menschen dazu, frei und bewusst zu entscheiden. Aus bildungstheoretischer Perspektive ist die Herstellung von Orientierung grundlegend vom Umgang mit Bestimmtheit und Unbestimmtheit geprägt (Marotzki 1990). Die Dichotomie zwischen Bestimmtheit und Unbestimmtheit hinsichtlich der Freiheitsfrage ist geprägt von widersprüchlichen und doch symbiotischen Zuständen: Der Mensch braucht einerseits Gewissheiten, um für sich selbst und bestimmt entscheiden zu können, um sich frei zu fühlen und

auch so zu handeln. Andererseits ist er von Zuständen der Unbestimmtheit abhängig, weil er *erstens* die komplexen, teils unsichtbaren Prozesse und Abläufe in seiner Umgebung nicht immer überblicken kann und damit darauf angewiesen ist, auf die jeweiligen Expert*innensysteme und das Wissen zu vertrauen. Die *zweite* Abhängigkeit von Unbestimmtheit ergibt sich durch das Prozesshafte des menschlichen In-der-Welt-Seins und der damit einhergehenden Befähigung, in unterschiedlichen Kontexten selbstbestimmt und frei entscheiden zu können. Wenn Menschen mit Situationen konfrontiert sind, in denen sie ihre bisherigen Routinen nicht mehr anwenden können, wenn damit ihre Weltaufordnungen[2], in Frage gestellt werden, also die Art, wie sie die Welt sehen, auf einmal fragwürdig erscheint (vgl. ebd., S. 153), dann können Transformationsprozesse des Selbst- und Weltverhältnisses angestoßen werden. In einer Welt voller Unübersichtlichkeiten und der permanenten Durchkreuzung von Ordnungsmustern ist die Herstellung individueller Gewissheiten und Routinen stets auch vom Umgang mit Unbestimmtheit geprägt. Der Umgang mit Unbestimmtheit ist damit auch ein wesentlicher Bestandteil, um sich neue Handlungsoptionen erst zu erschließen und somit Erfahrungen und Situationen neu zu kontextualisieren und erneut zu bewerten. Diese dynamische Wechselbeziehung zeichnet sich durch materielle und immaterielle Grenzen, fremd- oder selbstauferlegte Regeln und Gesetzmäßigkeiten aus, anhand derer Routinen entwickelt und validiert werden. Der Modus, in dem neue Handlungsspielräume herausgebildet werden, kann als ein *tentativer* gesehen werden. Jörissen und Marotzki (2009) beschreiben dies wie folgt:

„Unter den […] zeitdiagnostischen Bedingungen müssen wir mithin nach einem *anderen* Modus des Selbst- und Weltverstehens suchen, der mit Unbestimmtheit als solcher zurechtkommt. Die Struktur oder Logik des Selbst- und Weltverstehens muss dazu umgekehrt gedacht werden. Wenn im subsumptionslogischen Denken das Fremde unter bekannte Regeln und Schemata untergeordnet wird, *dann bildet das Allgemeine, bildet die Regel die Grundlage und den Ausgangspunkt.* Diese bleibt unverändert; alles Neue kann dann nur unter bestehende Regeln eingeordnet werden. Wenn es nicht passt, wird es sozusagen ‘passend gemacht’. Diese Struktur umzukehren würde bedeuten, das Neue, den Einzelfall, zur Grundlage zu machen und die passenden Regeln und Kategorien davon ausgehend erst zu *suchen.* In diesem Suchprozess wird mal das eine, mal das andere Verstehensmodell ausprobiert, bis man eines findet, mit dem man vorerst einigermaßen zurechtkommt – die *Relativität und Vorläufigkeit der eigenen Weltsicht* ist in diesem Modell von Anfang an enthalten. Wir bezeichnen diese Art des suchenden, immer unter dem Vorbehalt des „Als-ob“ agierenden Selbst- und Weltverhältnisses als *Tentativität.* Wir finden oder erfinden dabei Regeln, die für uns etwas zunächst unverständliches Neues zu etwas Verstehbarem machen. Die Regeln oder Schemata der Weltaufordnung sind dabei das, was verändert wird“ (ebd., S. 19, Hervorhebung im Original).

2 Mit dem Begriff der Welt*auf*ordnung werden alle von dem Menschen entwickelten kulturellen Schemata und Strukturen des Selbst- und Weltbild zusammengefasst, die es einem Menschen erlauben, seine Erfahrungen einzuordnen und zu bewerten.

Der Mensch als flexibles Wesen versucht also im Modus der Tentativität Bekanntes und Unbekanntes miteinander zu verbinden, Zusammenhänge für sich zu erklären und zu entdecken. Doch die Herstellung von Orientierung in der digitalen Welt ist paradox, weil digitale Technologien Orientierungsleistungen in unterschiedlichen Kontexten einfordern, obwohl insbesondere die Digitalen Medien auch dazu dienen, die materielle Welt zu organisieren. Menschen sind demnach permanent damit konfrontiert, sich in ihren Praktiken des Alltags, die immer undurchsichtiger und komplexer werden, mit „konstant verändernden materiellen Dingen" auseinanderzusetzen (Stalder 2016, S. 116). Es handelt sich dabei um Zusammenhänge, von denen man kaum wissen kann, woher sie kommen und was sie bedeuten. Dieses Phänomen lässt die Frage zu, wie Menschen es schaffen, sich in komplexer werdenden Strukturen zurechtzufinden und welche Strategien sie dazu ausbilden, denn wie Menschen Ordnungsstrukturen aufbauen, ist nicht immer deckungsgleich zu den algorithmischen Ordnungssystemen, die sie vorfinden. Es beschreibt vielmehr ein Wechselverhältnis von Ordnung und Durchkreuzung bestehender Strukturen, was dann im Zuge der Kontingenz zu kleineren oder größeren Irritationen führen kann. Die Herstellung von Orientierung ist deshalb von einer gewissen Komplexität geprägt, da die Prozesse zur Bewertung von Selbst- und Weltzusammenhängen in unterschiedlichen Kontexten und Relationen stattfinden können. Während die Sozialen Medien verschiedene Formen des Lifestyles und Weltansichten suggerieren, zu denen man sich verhalten kann, geben Suchergebnisse bereits ausgewählte Antworten auf spezifische Anfragen und strukturieren somit den Prozess der individuellen Entscheidungsfindung mit, beispielsweise dann, wenn man im Navigationsgerät eine Route plant oder man ein Restaurant in der Nähe sucht. In größeren Zusammenhängen werden Verhaltensweisen in der Nutzung und Inhalteproduktion durch Unternehmen erfasst, um so einzigartige, personalisierte Erlebnisse zu ermöglichen.

Doch die durch Algorithmen geprägten Filter- und Selektionsmechansimen bringen erst neue Unordnungen hervor. So lässt sich sowohl die Freiheit, eigene Ordnungen erstellen zu können festhalten, als auch die Verpflichtung, „eigene Ordnungen erstellen zu müssen, ob er [oder sie] auf diese Aufgabe nun vorbereitet ist oder nicht" (Stalder 2016, S. 116). Die große Unordnung bedingt sich durch die Frage, wie kulturelles Material geordnet werden muss, dass durch die digitalen Technologien nicht ausreichend gefiltert, organisiert und damit zugänglich gemacht wird, da mit Stalder (2016) gerade „das Übermaß sich oftmals widersprechender Bedeutungen, limitierter, regionaler, inkompatibler Ordnungen" zu „Unordnung und Bedeutungslosigkeit" führt (ebd., S. 116). Diese Bedeutungslosigkeit bedingt sich durch Reproduktion von kulturellen Gütern, wie beispielsweise der Erstellung von TikTok-Videos in gleichen Schemata wie gleicher Tonspur, gleiches Schnittmuster, vermeintlich gleiche Message, aber doch andere Personen und andere Kontexte. Auch wenn diese Reproduktion den eigenen Medienumgang schult und für neue eigene

Inhalte vorbereitet, weil man statt mit Vorlage, Zettel und Stift, nun mit anderem technischem Hilfsmittel ‚schreiben' lernt, werden Massen an Daten dieser digitalen Erprobungen und Übungen ins Netz gestellt und geteilt. Die medialen Praktiken auf den Plattformen der Sozialen Medien sind für die Diskussion um individuelle Freiheit von Relevanz, da sie sichtbar machen, wie Menschen ihre Freiheit und wo Teilhabe ermöglicht oder versperrt wird. Für die Betrachtung der medialen Praktiken sind aber auch die Strukturen relevant, ohne die es die digitalen Formate und medialen Artefakte nicht gäbe. Sie können daher nicht aus der Betrachtung ausgenommen werden, da in den digitalen Architekturen bereits Werte und Gesetzmäßigkeiten eingeschrieben sind, die einige Handlungsweisen gezielt ermöglichen und andere wiederum nicht.

Schon die Lockerung von Routinen können dabei Unbestimmtheiten erzeugen, manchmal auch Chaos, was dann wiederum zu Verunsicherungen führen kann. Marotzki (1990) beschreibt in seinem Entwurf einer strukturalen Bildungstheorie daher eine für Bildung notwendige flexible Haltung zu sich und zur Welt. Verbleibt der Mensch in seinen etablierten Routinen und damit in sich statisch, subsumiert er das Unbekannte in für ihn bekannte Deutungsweisen nur ungenügend, dann bleibt er womöglich unfrei. Die Bewegungen zur Herstellung von Bestimmtheit und der Umgang mit Unbestimmtheit beschreiben vor dem Hintergrund der Freiheitsfrage ein komplexes Spannungsverhältnis. Unfreiheit wird erst sichtbar und bewusst, wenn man erkennt, was man nicht sehen und verstehen kann. Gleichzeitig ist es mit Blick auf Unbestimmtheit auch erforderlich, sich bewusst zu machen, dass man nicht alles sehen und verstehen kann, um frei zu sein. Hieraus erwachsen verschiedene Paradoxien der Freiheit, die im Zusammenspiel mit Daten und den algorithmischen Berechnungen eine besondere Kontur bekommen.

4 Paradoxien der Freiheit

Die Konturierung von Freiheit im digitalen Zeitalter unterliegt mehreren Faktoren, die nicht immer transparent gemacht werden können. So kann der Einfluss von Algorithmen und Daten in täglichen und individuellen Handlungsvollzügen nicht immer rekonstruiert werden. Zugleich kann von der Handlung allein nicht auf die Willensbildung geschlossen werden. Algorithmen und ihre Infrastrukturen können als konstitutives Element für die Digitalität gesehen werden, sind jedoch in ihren Wirkweisen weitestgehend erst dann sichtbar, wenn Ergebnisse als Output dargestellt werden. Bevor es zur Darstellung von Ergebnissen und Ausgaben kommt, finden aber sehr viele Prozesse statt, die alle in ihren Abläufen für sich geregelt sind und teilweise ineinandergreifen.

Gerade durch das Zusammenspiel verschiedener Kontexte ergibt sich eine Komplexitätssteigerung, die strukturell bedingt ist. Freiheit wird beeinflusst durch die digitalen Technologien, die durch ihre Intransparenz als unbestimmbarer Möglichkeitsraum fungieren. Freiheit von Algorithmen und Daten meint dann, dass der Mensch in seinem Handeln innerhalb dieser Rahmenbedingungen dazu aufgefordert ist, sich zu den Gesetzmäßigkeiten zu verhalten, um Freiheitspotenziale zu entfalten. Dabei ist die Herstellung von Transparenz über die Abläufe und Implikationen der digitalen Technologien ein entscheidender Schritt, um schließlich erst bewusste und vernunftbegründete Entscheidungen zu treffen, Handlungen anzustoßen und damit auch die bestehenden Verhältnisse an sich zu hinterfragen. Eine weitere Perspektive, die wir auf Freiheit im digitalen Zeitalter werfen, beschreibt die Freiheit über Algorithmen und Daten. Das bedeutet für uns, dass durch die (bewusste) Nutzung von Algorithmen, Daten und die Gestaltung von digitalen Architekturen Potenziale von Freiheit zur Entfaltung kommen können.

Das hier konturierte Freiheitsverständnis ist von einer paradoxalen Prozesshaftigkeit geprägt, die wir als Paradoxien der Freiheit begreifen. *Paradox* ist Freiheit in der digitalen Welt deswegen, weil die Bedingungen für Freiheit gleichzeitig die Begründungen für die Einschränkungen von Freiheit im Hinblick auf Algorithmen und Daten sind. Deutlich wird dies beispielsweise, wenn über die Regulierung von digitalen Infrastrukturen, wie der Netzneutralität gesprochen wird oder wenn Inhalte in den Sozialen Medien zensiert beziehungsweise reguliert werden. Dabei geht es in erster Linie um das Spannungsverhältnis zur Sicherstellung von Freiheit und der Beschränkung eben dieser. *Prozesshaft* ist Freiheit grundlegend schon dadurch, dass sie immer wieder der Aushandlung unterliegt. Dies wurde bereits ausgeführt. In diesem Abschnitt sollen nun die zuvor dargestellten Bezugspunkte von Gesetzmäßigkeiten des Digitalen und den Implikationen für die Herstellung von Orientierung diskutiert werden. Dafür werden Paradoxien der Freiheit thematisiert, die sich entlang von Algorithmen und Daten aufzeigen lassen. Anhand der Paradoxien soll zudem danach gefragt werden, inwiefern der Freiheitsbegriff im Kontext von algorithmischen Strukturen selbst zur Disposition gestellt wird, wie wir entlang von Paradoxien um Individualisierung, Differenzierung und Orientierung aufzeigen werden.

4.1 Individualisierung

Die Individualisierung wird spätestens mit der Moderne zu einem zentralen Bezugspunkt von geistes-, sozial-, und erziehungswissenschaftlichen Auseinandersetzungen (Beck und Beck-Gernsheim 1994). Der Ausdifferenzierung von Lebensweisen, Arbeitsbedingungen und Märkten liegt die Wahlfreiheit zu Grunde. Es ist daher wenig überraschend, dass sich das Narrativ der Indivi-

dualisierung auch im Kontext der Digitalisierung und schließlich im Hinblick auf Digitalität manifestiert hat. Gleichzeitig verliert sich das Individuum in digital vernetzten Architekturen, dann, wenn es auf Berechnungen heruntergebrochen wird. Damit wird die Verfassung der individuellen Freiheit von der Heteronomie des Marktes und der digitalen Technologien überformt. Hieraus ergibt sich eine Paradoxie, die sich aus dem Streben nach Individualisierung einerseits und der Verwirklichung individueller Freiheit unter den Bedingungen der Heteronomie des Digitalen andererseits beschreiben lässt. Das Individuum wird zur Grundlage der Berechnung, es wird, mit Deleuze (1993) gesprochen zum Dividuum. Damit wird das vermeintlich Unteilbare in Datenpunkte unterschiedlicher Kontexte unter den Maßgaben der Kontrolle geteilt. Digitale Technologien sind heute mehr denn je „Kontrollformen mit freiheitlichem Aussehen“ (Deleuze 1993, S. 255) und dementsprechend im Spannungsfeld von Freiheit und Kontrolle zu verorten. Dies wird insbesondere im Hinblick auf Self-Tracking Apps deutlich, denn sie versprechen die Steigerung der individuellen Fitness und Leistungsfähigkeit. Ermöglicht wird dies nur durch die Kontrolle und Überwachung von Bewegungen mit unterschiedlichen Sensoren.

Für die Herstellung von Orientierung und die Konsequenzen für die individuelle Freiheit sind diese Phänomene in zweierlei Hinsicht bedeutsam: Erstens lassen sich über das Self-Tracking Routinen herausbilden, die dazu führen, dass Körper- und Selbstbeobachtungen in den Alltag integriert werden, indem Verhaltensweisen und Gewohnheiten bewusst, sichtbar und zählbar gemacht werden. Zweitens wird durch das Zählen der Daten ein Vergleichswert geschaffen, der einem über seinen Gesundheitszustand und Befindlichkeiten vermeintlich objektivierte Auskunft gibt. Einen Menschen in Daten zu erfassen, habe, so Damberger (2019), „etwas ausgesprochen Verführerisches“ (S. 307), da mit der Messbarkeit auch eine Möglichkeit der (Selbst-)Optimierung einhergeht. Man tritt also nicht nur zu anderen Menschen ins Verhältnis, sondern auch zu seinem früheren Ich, um vielleicht morgen besser zu sein als heute. Das Versprechen der individuellen Optimierung kann jedoch nur durch ein hohes Maß an Daten unterschiedlicher Qualität (Körper- und Bewegungsdaten, Ernährungs- und Verhaltensdaten) realisiert werden und steht dann auch im Zusammenhang mit den vielen anderen Menschen, deren Daten ebenfalls für Analysen und Auswertungen nutzbar gemacht werden. Die Methoden der Datensammlung und Auswertung durch die Unternehmen sind jedoch weitestgehend undurchsichtig. Zudem ist oftmals nicht klar, wie und unter welchen Bedingungen Dritte die Daten kontextualisieren können. Dies ist mindestens in zweierlei Hinsicht zu denken, während Unternehmen die Nutzungsdaten auswerten und damit Einblicke über die Geschäftsmodelle in andere Unternehmen bekommen, können Daten auch bewusst oder unbewusst von verschiedenen Akteurinnen und Akteuren geleakt werden. Während letzteres auf die Kontingenz im Zusammenhang zu Big Data und der Vernetzung allgemein abstellt und beispielhaft an Phänomenen zur Kontextualisierung von

großen Datenmengen herangezogen werden kann (vgl. Verständig 2020, S. 129), lässt sich der Kontroll- und Überwachungseinfluss durch Unternehmen beispielhaft darüber veranschaulichen, wenn Mailprovider die Rechnungen der Nutzer*innen auf das Kaufverhalten hin untersucht, um ihnen zielgruppenspezifische und hochgradig individualisierte Werbung zu schalten.

Deleuze (1993) hält in seinem Postskriptum über die Kontrollgesellschaften schließlich fest, dass weder zur Furcht noch zur Hoffnung Grund bestehe, „sondern nur dazu, neue Waffen zu suchen“ (ebd., S. 256). Im Horizont der Digitalität sind diese neuen Waffen dann die Algorithmen und Daten, die zu strukturellen Veränderungsprozessen führen und diese gleichzeitig mitprägen. Während der Mensch sich individuell entfaltet, tritt er im Hinblick auf statistische Berechnungen und die marktwirtschaftliche Einbettung seines Handelns in den Hintergrund und wird vielmehr als berechenbarer Bestandteil gesellschaftlicher Strukturen gesehen. O’Neil (2016) beschreibt aus mathematischer Sicht die sozialen Implikationen von Algorithmen und die Konsequenzen für verschiedene Felder, wie Bildung, Beruf und Privatleben. Die von O‘Neil thematisierten Algorithmen und die zugrunde liegenden mathematischen Modelle werden als „Weapons of Math Destruction“ gefasst, wobei ein besonderer Fokus auf die Kraft der Berechnung einerseits und den Glauben in diese Berechnungen andererseits gelegt wird. Die algorithmischen Systeme sind undurchsichtig, skalierbar und richten Schaden entkoppelt von ihren aktuellen Abläufen an, indem sie die scheinbar sozio-ökonomisch Schwachen bestrafen und die die Menschen belohnen, die ohnehin von diesen Systemen profitieren. Die technologischen Entwicklungen führen damit zu einer Verstärkung von bestehenden Ungleichheiten, wie auch Eubanks (2018) aufgezeigt hat.

4.2 Differenzierung

Digitalität ist von Algorithmizität geprägt und algorithmische Strukturen berechnen und unterscheiden. Algorithmen differenzieren. Differenzierungen sind dann die Grundlage dafür, ob jemand das eine oder das andere Produkt beim Shopping angezeigt bekommt, ob man in die eine, die andere oder gar keine Gehaltsgruppe eingestuft wird. So findet beispielsweise in der Arbeitswelt eine auf Algorithmen und Daten basierte Differenzierung über die Selektion von Stellenbewerber*innen oder durch die Bestimmung unterschiedlicher Entgelte und Arbeitsanforderungen und -Bedingungen statt, um so die höchstmögliche Passung für Organisationen und Mitarbeitende zu schaffen. Differenzierungen können auch aus sozialen Gründen erfolgen und beispielsweise Faktoren wie Bedürftigkeit, Chancenverteilung und Teilhabe berücksichtigen. Modi der Differenzierung sind vor dem Hintergrund von Algorithmen und Daten dahingehend von einer Paradoxie geprägt, da die Unterscheidung des Einzelnen auf statistischen Berechnungen der Masse beruht,

dementsprechend haben individuelle Verhaltensmuster einer Person im Rahmen einer statistischen Kontextualisierung auch Auswirkungen auf andere. Es handelt sich gewissermaßen um eine Produktion von Unwahrscheinlichkeit durch Kalkül (vgl. Esposito 2014, S. 237), weshalb die Differenzierung insbesondere im Hinblick auf die Herstellung von Orientierung bedeutsam ist.

Durch den von Zuboff (2019) gefassten Überwachungskapitalismus tritt diese Problemlage besonders deutlich hervor. Menschliches Verhalten werde in berechenbare und datenbasierte Mechanismen gedrängt. Diese „dehumanisierenden“ Methoden sowie die daraus resultierende „radikale Indifferenz“, stehen der freien Entfaltung, dem freien Willen entgegen und untergraben die individuelle Selbstbestimmung (ebd., S. 345 f.). Eine für die Betrachtung von Freiheit notwendige Voraussetzung ist dafür jedoch die Transparenz um die Wirkweisen von algorithmischen Strukturen. Freiheit wird zumindest dann zur Bedingung, wenn es um die individuelle Teilhabe über und mit digitalen Medien geht. Die individuelle Freiheit steht historisch geprägt im Spannungsverhältnis zu Gesetzen und gesellschaftlichen Normen. Wenn Algorithmen und digitale Daten nun die Art und Weise beeinflussen, wie wir Menschen die Welt und uns sehen, dann müssen Aushandlungen über die individuelle Freiheit auch konsequent entlang der digitalen Technologien verlängert werden. Es muss öffentlich diskutiert werden, wie und inwieweit digitale Daten uns als Menschen beschreiben können und sollen. Es muss aber auch in die andere Richtung gefragt werden, wie wir uns als Menschen durch digitale Technologien weiterentwickeln können. Transparenz meint dann die Dokumentation von Datensätzen und Modellen, den Zugang zu Informationen über die Wirkweisen von Systemen von gesellschaftlichem und öffentlichem Interesse sowie die Aufbereitung der Informationen, um sie der Allgemeinheit tatsächlich zugänglich zu machen.

Das hier dargelegte Freiheitsverständnis steht im engen Zusammenhang zum Besitzanspruch von Daten, doch der Besitz von Daten ist hochgradig problematisch, denn er unterliegt anderen Gesetzmäßigkeiten, als der Besitz von materiellen Gütern und gleichzeitig sind die für die Interaktionen zugrunde liegenden Daten in verschiedene Prozesse und oftmals unterschiedliche Kontexte eingebettet. Während die durch das Self-Tracking erhobenen Gesundheitsdaten ein Hilfsmittel zur individuellen Verbesserung der Gesundheit dienen können, sind sie zugleich auch ein Dokument für die täglichen Routinen, die tief in die Privatsphäre der Menschen hineinreichen. Doch Transparenz allein genügt nicht, um individuelle Freiheit durch Differenzierung zu ermöglichen. Denn die Differenzierungsleistungen erfolgen weitestgehend fremdbestimmt auf Seite der Unternehmen und Organisationen, die Unterscheidungen vornehmen.

Differenzierungen sind nicht zuletzt auch die Grundlage für Ungleichbehandlungen von Menschen und stellen damit die Schwelle zu Diskriminierungen dar. Das muss nicht zwangsläufig heißen, dass Ungleichbehandlungen

auch Diskriminierungen sind, jedoch ist die Sensibilität hinsichtlich der Grenze und ihrer Überschreitung durch algorithmische Systeme besonders mit Blick auf lernende Systeme und Künstliche Intelligenz wichtig (Zweig 2019).

4.3 Diskriminierung

Wenn Differenzierungen zu Diskriminierungen führen, dann meist, weil algorithmische Systeme entgegen ihrer Intention soziale Schieflagen reproduzieren oder durch die Werte, die ihnen eingeschrieben sind, soziale beziehungsweise kulturelle Zusammenhänge nicht berücksichtigen (Apprich et al. 2018). Diese Form der Diskriminierung wird dann strukturell über digitale Technologien abgebildet, was allein deshalb schwierig ist, weil die statistische Berechnung eben nicht auf die Intention der Entwickler*innen, der Daten oder der Software selbst zurückzuführen ist. Unabhängig von der Feststellung von Diskriminierungsweisen im Zusammenhang von Algorithmen und Daten bleibt für die individuelle Freiheit auch festzuhalten, dass die Minimierung von Klassifikatoren der Diskriminierung in digitalen Technologien vermutlich nur eine Verschiebung von Mechanismen zur Reproduktion von sozialen Ungleichheiten in anderen gesellschaftlichen Feldern zur Folge haben wird (vgl. Eubanks 2018, S. 143 ff.).

Zudem ist die Kontextualisierung von Daten besonders dann relevant, wenn es um diskriminierungsfreie algorithmische Systeme geht. Liegen beispielsweise schon Verzerrungen bei Trainingsdaten von lernenden Systemen vor, dann werden diese Schieflagen in die Berechnungen übernommen und damit die Ergebnisse ebenfalls verzerrt. Zugleich können Missinterpretationen oder fehlende Parameter dazu führen, dass Menschen aus unterschiedlichen sozialen Gruppierungen oder mit bestimmten Merkmalen strukturell nicht berücksichtigt und dadurch diskriminiert werden. Buolamwini (2017) hat beispielsweise über die Auswirkungen von Fehlklassifizierungen und die Bedeutung des Aufbaus umfassender Trainingssätze sowie Benchmarks wertvolle Einsichten in genderspezifische Diskriminierungen durch algorithmische Systeme der Mustererkennung herausgearbeitet. Da im Prototyp einer Gesichtserkennungssoftware vorwiegend weiße Gesichter als Trainingsdaten dienten, hat die Software schwarze Frauen schlechter erkannt. Die hier gesellschaftsrelevanten algorithmischen Strukturen sind datengetrieben, sie betreffen beispielsweise auch die Risikoeinschätzung von bestimmten Personengruppen. So beschreibt Eubanks (2018) unter anderem, wie die Los Angeles Homeless Services Authority (LAHSA) ein System entwickelt und eingesetzt hat, welches mithilfe von erhobenen Daten und einem für diesen Zweck entwickelten Algorithmus Wohnraum für Obdachlose finden und so die Wohnungsvermittlung erleichtern soll (vgl. ebd., S. 70 ff.). Dies ist jedoch nur dann wirklich zielführend, wenn auch tatsächlich ein ausreichender Pool von Wohnungen zu vergeben ist, an-

dernfalls werden Selektionsmechanismen umgesetzt, die Menschen in ihrer Not kategorisiert und entsprechend priorisiert. Zuboff (2019) macht deutlich, wie sich neben der politischen Überwachung eine Form der ökonomischen Überwachung am Markt durchgesetzt hat, die durch Automatisierung und die Bearbeitung von großen Datenmengen zwar den kleinsten gemeinsamen Nenner als Menschen definiert, jedoch das Menschliche aus den sozialen Interaktionen durch die blinde Berechnung gänzlich heraus dividiert.

Für die Diskussion um Freiheit sind diese Herausforderungen und Problemfelder besonders relevant, da sie die ethischen Dimensionen der technologischen Entwicklungen berühren. Wenn die Gesetzmäßigkeiten des Digitalen also nicht nur durch Algorithmen, als den Menschen vorgelagerte und vermeintlich neutrale Objekte gesehen werden können, sondern in ihrer Interaktion und Beschaffenheit zum Menschen gesehen werden, dann muss der Blick auch auf die Daten gerichtet werden, mit denen Menschen und Algorithmen arbeiten. Die Daten, die dem Gebrauch und der Verarbeitung zugrunde liegen und aus den Interaktionen hervorgehen, sind damit gewissermaßen Ressourcen, die aber einer gänzlich anderen Logik folgen als natürliche Rohstoffe. Gemeinhin wird auch davon gesprochen, dass Daten das neue Öl sind, doch diese Analogie greift zu kurz[3]. Denn Öl als Rohstoff bleibt immer Öl, es wird immer fließen und eine bestimmte Viskosität haben. Daten sind in ihrer Beschaffenheit hochgradig heterogen und müssen längst nicht zur Herstellung von Informationen führen und genau darum ist die Masse an Daten ein wichtiger Faktor. Wir produzieren daher immer mehr Daten, die unsere Umwelt und uns selbst beschreiben und dabei häufig auch in Unordnung und Bedeutungslosigkeit verschwinden. Während der Abbau von mineralen Rohstoffen unheimlich viel kostet, ist die Datenerfassung, die Sammlung von Verhaltensdaten oder die Überwachung verhältnismäßig kosteneffektiv, da die digitalen Infrastrukturen dafür bereits etabliert sind. Der Aufwand liegt in der Kontextualisierung, Auswertung und Interpretation von Daten. Daten können damit neue Möglichkeiten der Welterkundung und der Wissensgenerierung eröffnen, es hängt dabei jedoch von den Menschen ab, wie sie mit den Daten umgehen, um nicht blind oder ohnmächtig zu sein, sondern Orientierung herzustellen.

5 Fazit

Der Beitrag hat seinen Ausgang in der Frage gehabt, inwiefern Freiheit im digitalen Zeitalter konturiert werden kann. Dabei wurde ein weites Konzept von Freiheit umrissen, um interdisziplinäre Anknüpfungspunkte zur Auseinandersetzung mit algorithmischen Strukturen zu ermöglichen. Mit Höffe (2018)

3 Diese Formulierung lässt sich auf einen 2017 publizierten Beitrag im Economist zurückführen.

wurde gezeigt, wie Freiheit zeitgemäß in unterschiedlichen Themenbereichen gedacht werden kann und welche strukturellen Bezugspunkte es gibt. Damit wurde auch aufgezeigt, dass Freiheit kein statischer Zustand, sondern vielmehr als dynamischer Prozess geprägt von Aushandlungen und verschiedenen Spannungsfeldern ist. Darauf aufbauend wurden Digitalität, die Gesetzmäßigkeiten des Digitalen und die individuelle Herstellung von Orientierung verhandelt, um so die paradoxalen Strukturen von Freiheit im digitalen Zeitalter entlang von Individualisierung, Differenzierung und Diskriminierung zu beschreiben.

Will man den kantischen Gedanken zum Ausgang aus der selbstverschuldeten Unmündigkeit hier aufleben lassen, dann sind den digitalen Technologien große Freiheitspotenziale eingeschrieben. Durch die Berechnungen von großen Datensätzen können wir heute mehr denn je über die Welt erfahren, in der wir leben. Es werden strukturelle Probleme sichtbar und vor allem belegbar, auch wenn die Möglichkeit einer Verstärkung dieser sozialen Schieflagen besteht. Sie sind nicht nur das sichtbare Ergebnis systematischer, fehlerhafter Berechnungen, sie erinnern uns auch daran, dass ein fehlerhaftes Gesellschaftssystem nicht durch den Einsatz von Technologien repariert werden kann. Den aufgezeigten Paradoxien wohnen hohe Bildungspotenziale inne. Dies zeigt sich einerseits in der Möglichkeit zur Reflexion über Algorithmen und Daten und andererseits in der Dekonstruktion der bestehenden Missstände. Unsichtbare algorithmische Infrastrukturen machen zunächst unsichtbare strukturelle Probleme sichtbar – allerdings müssen wir lernen, auch hinzusehen und vor allem planvoll und verantwortungsbewusst zu handeln. Algorithmische Systeme bilden Ordnungen heraus und sie lassen damit Orientierungsrahmen zu, gleichzeitig können automatisierte Prozesse der Filterung, Selektion und Repräsentation von Daten aufgrund ihrer ökonomischen Einbettungen und ihrer politischen sowie kulturellen Implikationen nur begrenzt als Katalysator für Freiheitsbestrebungen gesehen werden, denn die Algorithmen, die den Geschäftsmodellen der Dienste zugrunde liegen, sind oftmals nicht transparent und ihre Wirkweisen dadurch häufig nur schwer rekonstruierbar.

Algorithmen repräsentieren nicht nur die Welt in der wir leben, sie prägen unsere Welt und stehen im engen Wechselverhältnis zu menschlichen Handlungen und Entscheidungen. Daten sind dann nicht das objektive Resultat von vermeintlich neutralen Berechnungen, sie sind hochgradig performativ und von sozialen sowie kulturellen Einflüssen geprägt und damit keineswegs neutral. Die individuelle Freiheit wird mit fortschreitender Technologisierung zu einem Faktor der Berechnung, der jedoch mehr Unbestimmtheit als Bestimmtheit hervorbringt. Ohne Algorithmen sind wir blind. Ohne das Wissen um Algorithmen und ihre meist impliziten Wirkweisen sind wir allerdings auch nicht frei.

Literatur

Apprich, C., Chun, W. H. K., Cramer, F., & Steyerl, H. (2018). *Pattern Discrimination.* University of Minnesota Press.

Beck, U., & Beck-Gernsheim, E. (Hrsg.). (1994). *Riskante Freiheiten: Individualisierung in modernen Gesellschaften* (1. Aufl., Erstausg). Suhrkamp.

Bridle, J. (2018). *New dark age: Technology, knowledge and the end of the future.* Verso.

Buolamwini, J. A. (2017). *Gender Shades: Intersectional Phenotypic and Demographic* [Master's Thesis, Massachusetts Institute of Technology]. https://dam-prod.media.mit.edu/x/2018/02/05/buolamwini-ms-17_WtMjoGY.pdf

Damberger, T. (2015). *Quantified Self and Transhumanism – A Means to an End* (J. Hughes, Hrsg.). http://ieet.org/index.php/IEET/more/damberger20150817

Damberger, T. (2019). *Bildung im Digitalzeitalter: Zur pädagogisch-anthropologischen, technischen und medienpädagogischen Dimension des Verhältnisses von Bildung und Digitalisierung* [Habilitation, Otto-von-Guericke-Universität Magdeburg]. https://opendata.uni-halle.de//handle/1981185920/32109

Daub, A. (2020). *Was das Valley denken nennt über die Ideologie der Techbranche.* Suhrkamp.

Deleuze, G. (1993). Postskriptum über die Kontrollgesellschaften. In G. Deleuze (Hrsg.), *Unterhandlungen 1972–1990.* Suhrkamp, S. 254–262.

Esposito, E. (2014). Algorithmische Kontingenz. Der Umgang mit Unsicherheit im Web. In A. Cevolini (Hrsg.), *Die Ordnung des Kontingenten,* Springer Fachmedien Wiesbaden, S. 233–249. https://doi.org/10.1007/978-3-531-19235-2_10

Grünberger, N. (2021). Postkolonial post-digital: Forschungsfelder und Anschlussstellen für die Medienpädagogik durch eine postkoloniale Perspektive auf eine Post-Digitalität. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, 211–229. https://doi.org/10.21240/mpaed/jb16/2021.02.25

Habermas, J. (2021). Überlegungen und Hypothesen zu einem erneuten Strukturwandel der politischen Öffentlichkeit. In M. Seeliger & S. Sevignani (Hrsg.), *Ein neuer Strukturwandel der Öffentlichkeit?.* Nomos, S. 470–500. https://doi.org/10.5771/9783748912187-470

Kühnlein, M. (Hrsg.). (2018). *Das Risiko der Freiheit: Im interdisziplinären Gespräch mit Otfried Höffe* (1. Auflage). Nomos.

Lessig, L. (1999). *Code and other laws of cyberspace.* Basic Books.

Marotzki, W. (1990). *Entwurf einer strukturalen Bildungstheorie: Biographietheoretische Auslegung von Bildungsprozessen in hochkomplexen Gesellschaften: Zugl.: Hamburg, Univ., Habil.-Schr., 1989* (Bd. 3). Dt. Studien-Verl.

Matzner, T. (2017). Opening Black Boxes Is Not Enough – Data-based Surveillance In Discipline and Punish And Today. *Foucault Studies*, *0*(23), 27–45. https://doi.org/10.22439/fs.v0i0.5340

Matzner, T. (2019). The Human Is Dead – Long Live the Algorithm! Human-Algorithmic Ensembles and Liberal Subjectivity. *Theory, Culture & Society*, *36*(2), 123–144. https://doi.org/10.1177/0263276418818877

Noller, J. (2021). Philosophie der Digitalität. In U. Hauck-Thum & J. Noller (Hrsg.), *Was ist Digitalität?.* Springer Berlin Heidelberg, S. 39–54. https://doi.org/10.1007/978-3-662-62989-5_4

O'Connell, M. (2017). *To be a machine: Adventures among cyborgs, utopians, hackers, and the futurists solving the modest problem of death* (First edition). Doubleday.

Pasquale, F. (2015). *The Black box society: The secret algorithms that control money and information* (First Harvard University Press paperback edition). Harvard University Press.

Reckwitz, A. (2017). *Die Gesellschaft der Singularitäten: Zum Strukturwandel der Moderne.* Suhrkamp.

Rosenblat, A. (2018). *Uberland: How algorithms are rewriting the rules of work.* University of California Press.

Schenk, S., & Karcher, M. (Hrsg.). (2018). *Überschreitungslogiken und die Grenzen des Humanen: (Neuro-)Enhancement – Kybernetik – Transhumanismus.* Martin-Luther-Universität.

Stalder, F. (2016). *Kultur der Digitalität* (Originalausgabe, erste Auflage, Bd. 2679). Suhrkamp.

Stalder, F. (2021). Was ist Digitalität? In U. Hauck-Thum & J. Noller (Hrsg.), *Was ist Digitalität?.* Springer VS, S. 3–7. https://doi.org/10.1007/978-3-662-62989-5_1

Zweig, K. A. (2019). *Ein Algorithmus hat kein Taktgefühl: Wo künstliche Intelligenz sich irrt, warum uns das betrifft und was wir dagegen tun können* (Originalausgabe). Heyne.

Geist in der Maschine

Über Bildung, Schein und Wahrheit im Digitalzeitalter

Thomas Damberger

1 Einleitung oder: Dystopie der Maschine

Im Jahr 1909 legt der britische Autor Edward Morgan Forster seine Erzählung *Die Maschine steht still* (im Original: *The Machine Stops*) vor. Forster beschreibt darin eine nicht näher bestimmte Zukunft, in der Menschen nicht mehr auf der Erde, sondern alleine, je für sich, in unterirdischen wabenförmigen Zimmern leben. Die Kommunikation mit anderen Menschen findet mithilfe einer Maschine statt, die alles mit allem verbindet. Den Zustand, den Forster skizziert, könnte man treffend als ein permanentes Online-Sein bezeichnen. Aber es ist weitaus mehr, was die Maschine für die Menschen bedeutet. Sie versorgt sie mit Nahrung, Musik und frischer Luft, sie lässt auf Knopfdruck Bett und Badewanne erscheinen, sie macht es nicht erforderlich, sich vom Stuhl zu erheben, was schlussendlich dazu führt, dass der menschliche Körper nur noch von marginaler Bedeutung ist. So wird Vashti, eine der Protagonistinnen der Erzählung, als ein haarloser, zahnloser „in Tücher gewickelter Fleischberg" (Forster 2016, S. 9) beschrieben.

Vashti hat abertausende Bekannte, kennt aber niemanden persönlich, im Grunde noch nicht einmal ihre eigenen Kinder. Die Maschine sorgt dafür, wer sich mit wem fortpflanzt. Über den Fortpflanzungsakt hinaus, bei dem unklar bleibt, ob es sich um ein körperliches oder technisches Geschehen handelt, findet keine unmittelbare zwischenmenschliche Beziehung statt. Direkt im Anschluss an die Geburt werden Mutter und Kind getrennt. Das Kind wächst in einer Erziehungsanstalt auf und erhält ab einer bestimmten Reife eine eigene Wabe. Der Tod indessen wird ebenfalls nicht dem Zufall überlassen. Man stirbt in der Regel durch Euthanasie, und es ist die Maschine, die bestimmt, ob und wann man ein Anrecht darauf hat, in dieser Weise aus dem Leben zu scheiden (vgl. ebd., S. 36).

Als erstrebenswert gilt in Forsters Erzählung nicht das unmittelbare Erleben, sondern das, was als Destillat vieler Geister charakterisiert wird. Gemeint sind Vorträge und besser noch Metavorträge, die man vermittelt über die Maschine, über Grammophon und Kinematophoten, sehen und hören kann, ohne dass eine primäre Erfahrung vonnöten ist. Der Mensch erscheint hier als Medium und reinigender Filter, die Maschine hingegen als Metafilter. Das Destillat am Ende des Reinigungsprozesses ist die Idee, und das Sammeln von Ideen erweist sich als der Sinn der Existenz in Forsters Dystopie: „‚Hütet euch vor Ideen aus erster Hand!‘, proklamierte einer der fortschrittlichsten Vertreter. ‚Es gibt im Grunde keine Ideen aus erster Hand. Es handelt sich um bloße Reaktionen des Körpers auf Liebe und Angst, und wer könnte auf einem so rohen Fundament schon eine Weltanschauung errichten? Bezieht eure Ideen aus zweiter Hand, wenn möglich aus zehnter Hand, erst dann ist das verstörende Moment des direkten Erlebens ausreichend abgeschwächt. Erarbeitet euch mein Fachgebiet, die Französische Revolution, nicht im Alleingang! Erarbeitet euch stattdessen, was ich denke, was Enicharmon dachte, was Urizen dachte, was Gutch dachte, was Ho-Yung dachte, was Chi-Bo-Sing dachte, was Lafcadio Hearn dachte, was Carlyle dachte, was Mirabeau über die Französische Revolution gesagt hat. Diese zehn Geistesgrößen destillieren aus dem vergossenen Blut von Paris und den eingeschlagenen Fenstern von Versailles eine Idee, die euch im Alltag von größtem Nutzen sein wird“ (ebd., S. 58 f.).

Das Dystopische kommt dabei nicht zuletzt durch den Umstand zur Geltung, dass die Maschine und ihre Funktionsweise – weniger noch ihre Macht – von niemanden mehr verstanden wird. Zwar ist bekannt, dass sie irgendwann einmal von Menschenhand geschaffen wurde, aber mehr als ein Handbuch (das einzige vorhandene Buch, das paradoxerweise nie zur Hand genommen wird) existiert nicht.

Zunehmende Störungen in der Maschine, die anfangs vernachlässigt werden, jedoch immer mehr zunehmen und zuletzt das Leben bedrohen, nötigen die körperlich degenerierten Menschen zum Verlassen ihrer Zimmer. Die mit giftigen Gasen durchsetzte Luft der zerstörten Erdoberfläche kann von der Maschine nicht mehr gereinigt werden, was für die Menschen letztlich den sicheren Tod bedeutet.

Die von Forster beschriebene Maschine ist Ausdruck eines außer Kontrolle geratenen, umfassenden kybernetischen Systems, d. h. eines Systems, das auf Steuerung und Kontrolle basiert, sich verselbstständigt hat und für seinen Erbauer zuletzt existenzbedrohend wird. Sie skizziert zudem in groben Zügen das, was wir mit dem Internet und mehr noch, dem Internet der Dinge, derzeit zum Teil erleben und teils als mögliche Gefahr diskutieren. Angesprochen ist beispielsweise die Frage nach dem Wert und der Bedeutung medial vermittelter Erfahrungen. Nicht zuletzt ist die Maschine Ausdruck einer perfektionierten Naturbeherrschung, die an den jeweils äußersten Begrenzungen menschlicher Existenz beginnt bzw. endet. Innerhalb dieser Begrenzung wird jede mensch-

liche Regung diszipliniert, werden Eigensinn und Spontaneität maximal reduziert und eine gleichsam reduzierte, instrumentelle Vernunft auf alle Lebensbereiche ausgedehnt. Die Maschine ist damit auch Ausdruck einer spezifischen Pädagogik des 18. bzw. 19. Jahrhunderts und dem mit ihr einhergehenden Verhältnis zur Natur (vgl. Berr 1994, S. 79).

Ein Blick auf die historische Genese der Maschine (2) lässt die genannten Aspekte aufleuchten und eröffnet ein Verständnis für die Weise, wie das Maschinenhafte in Gestalt einer von Algorithmen durchdrungenen, datafizierten Welt wirkt (3). Offenkundig wird dabei eine zunehmende Entgrenzung, die sich gegenwärtig als symptomatisch für das Verhältnis von Mensch und Maschine erweist (4). Eine mögliche Konsequenz dieser Entgrenzung ist eine Rekursivität, die auf eine Lücke verweist, die mit dem Menschsein einhergeht und für die Bildung des Menschen zum Menschen von höchster Relevanz ist (5). Auf diese Lücke hinzuweisen ist der Sinn und Zweck dieses Beitrags.

2 Geschichte der Maschine

Die Begriff Maschine geht auf den altgriechischen Ausdruck mēchanḗ (μηχανή) zurück und bedeutet Mittel oder List. Aristoteles (385–323 v. Chr.) sieht im Maschinenhaften das Gegenteil der Natur. Die Natur bzw. phýsis (φύσις) ist für ihn eine materielle Substanz, in die eine Bewegung hineingelegt wurde, womit gemeint ist, dass weder die Substanz noch ihre inhärente Bewegung künstlich erzeugt werden (vgl. Aristotle 2008, S. 33); dies gilt im Übrigen auch für die Ruhe, die als die andere Seite der Bewegung zu verstehen ist. Das Blau des Meeres ruht in der Bewegung, und im Baum, der am selben Ort bleibt und im Sinne dieser Ortsbeständigkeit ruht, offenbart sich die Bewegung als Veränderung im Zuge der Jahreszeiten. Bewegung bedeutet zugleich auch Umschlag im Sinne eines Aus- bzw. Durchschlagens. Das, was zunächst als bloße Möglichkeit im Menschen vorhanden ist, vermag in einer bestimmten Situation, in einer eigentümlichen Stimmung auszuschlagen, wird sichtbar, gerät in die Unverborgenheit (vgl. Heidegger 2004 [1939], S. 249). Die Bewegung betrifft daher wesentlich (auch) das Verhältnis von Potenzialität und Realität.

Im Gegensatz dazu steht für Aristoteles das Maschinenhafte bzw. Mechanische. Ihr ist die Bewegung keineswegs inhärent – sie ist vielmehr von außen initiiert. Das Maschinenhafte zeugt für Aristoteles zwar von Kunstfertigkeit, trägt aber nicht zur wissenschaftlichen Erkenntnis der Natur bei. Dieser Gedanke mag insbesondere aus unserer heutigen Perspektive zunächst irritieren, nutzen wir doch gegenwärtig Maschinen beispielsweise in Form von Fitnessarmbändern oder Smartwatches, um via *Selftracking* Informationen über unser

Bewegungsprofil und diverse Körperfunktionen zu erhalten. Tatsächlich wird aus Aristoteles' Sicht im Zuge einer maschinengestützten Erkenntnis nichts Neues erkannt, sondern lediglich das, was zuvor schon erkannt worden ist und als prinzipiell erscann- und datafizierbar bestimmt wurde, hervorgekehrt. Die Maschine dient also dazu, dasjenige, was bereits grundsätzlich als festgestellt gilt, in den Bestand zu überführen und jederzeit bestellbar, d. h. auswertbar zu machen. Sie hat in diesem Sinne keine entbergende Funktion, d. h. sie dient nicht dazu, etwas Verborgenes in die Unverborgenheit zu bringen. Dies würde ein spezifisches, offenes, seinlassendes, wartendes (nicht *erwartendes*) Verhältnis voraussetzen.[1] Vielmehr ist sie Ausdruck des Machbaren und damit ein Moment menschlicher Macht, die sich selbst nicht zu überschreiten vermag und damit notwendigerweise innerhalb dessen, was zuvor als machbar bestimmt wurde, verbleibt.

Für den Philosophen und Mathematiker René Descartes (1596–1650) sind es im Kern drei Aspekte, die den Unterschied zwischen Mensch und Maschine markieren. Zum einen handelt es sich um den reflektierten Einsatz von Sprache, zum anderen um das Vernunftvermögen und zuletzt um den Zugang zur denkenden und ausgedehnten Substanz (*res cogitans* und *res extensa*). Die Vernunft, um die es Descartes geht, ist eine instrumentelle. Sie dient einerseits der Selbstvergewisserung: „[J]e pense, donc je suis" (Descartes 1960 [1641], S. 53) und damit der Selbstbemächtigung, um davon ausgehend ein Wissen von der Welt als sichere Grundlage, d. h. mithilfe wissenschaftlicher Methode, generieren zu können (vgl. ebd., S. 53 ff.). Andererseits erweist sich die Vernunft als das Vermögen, den Menschen als körperliches und geistiges Wesen zu begreifen. Wie allerdings mentale Zustände auf den Körper, der von Descartes mechanistisch verstanden wurde, einwirken können, wird im wissenschaftlichen Kontext bis in die Gegenwart hinein im Rahmen des sogenannten Leib-Seele-Problems diskutiert (Beckermann 2011, Hermanni & Buchheim 2006, Ravenscroft 2008). Descartes geht davon aus, dass der Geist den gesamten Körper durchdringt, aber mithilfe der Zirbeldrüse (corpus pineale), die aus seiner Sicht für die Koordination der Gehirnzustände als verantwortlich gilt, eine Verbindung mit der *res extensa* herstellen kann (vgl. Descartes 1870, S. 112).

Diesen Dualismus überwindet der Arzt Julien Offray de La Mettrie mit seinem 1748 zunächst anonym veröffentlichten Werk *Maschine Mensch* (im Original: *L'Homme-Machine*): „Descartes und alle Cartesianer […] haben denselben Fehler gemacht. Sie haben zwei genau zu unterscheidende Substanzen in dem Menschen angenommen, als ob sie solche gesehen und richtig gezählt hätten" (La Mettrie 2016 [1748], S. 39). La Mettrie vertritt eine monistische Position, dergemäß der Mensch in seiner Gänze mechanistisch erklärt

1 An dieser Stelle sei auf die technikphilosophischen Überlegungen Heideggers verwiesen, in deren Sinne das Bestellen, das Ent- und Verbergen, das Unverborgene, das Seinlassen und das (Er-)Warten in der angeführten Passage zu verstehen sind (vgl. Heidegger 1982 [1962]).

werden kann. Es handelt sich hierbei um eine spezifische Form des Materialismus – und zwar des mechanistischen Materialismus. Ausgangspunkt ist dabei die Materialität, aus der alle weiteren Zustände resultieren und zumeist mit naturwissenschaftlichen Methoden erforscht werden. Insofern Letzteres geschieht, spricht man auch vom Naturalismus bzw. naturalistischen Materialismus: „Naturalistische Materialisten schätzen vor allem den Erkenntniswert der Naturwissenschaften, der als einzig sicher erachtet wird. […] Der Naturalismus *naturalisiert* verschiedene Bereiche des Menschlichen, d.h. Sprache, Erkenntnis, Psyche, Soziales werden als Spezialfälle der physikalischen Natur angesehen und sollen mithilfe der Methoden der Naturwissenschaften erforscht werden" (Küpper 2017, S. 115f., Hervorh. im Original). Die hier angesprochenen Spezialfälle werden als *emergente* Phänomene verstanden (vgl. Bunge & Mahner 2004, S. 148). Bewusstsein ist demzufolge immer ein Zustand von Dingen; ein verändertes Bewusstsein (Träumen etc.) ist eine Zustandsveränderung von Dingen, und der Tod gilt als das irreversible Nicht-mehr-Funktionieren (z.B.) des menschlichen Körpers als Ding.

Der schon erwähnte Martin Heidegger sieht insbesondere in der modernen Maschine ein Ausdruck des Wesens moderner Technik. Diese unterscheide sich von der klassischen Technik dadurch, dass sie keinen hervorbringenden, sondern ausschließlich formenden und bearbeitbaren Charakter habe. Um diesen Gedanken besser nachvollziehen zu können, macht es Sinn, zunächst einmal das Wesen der klassischen Technik zu skizzieren. Dieses Wesenhafte fasst Aristoteles in seiner Ursachenlehre zusammen; er spricht hier von insgesamt vier Ursachen: dem Material (causa materialis), der Form (causa formalis), dem Zweck bzw. dem Ziel (causa finalis) und dem Bewirken (causa efficiens) (vgl. Höffe 2006, S. 117). Die Bildhauerin schlägt aus dem Marmorblock (Material) die Figur eines Löwen. Sie hat eine Vorstellung davon, wie der Löwe aussieht, welche Form er hat. Sie setzt die Werkzeuge an und verfolgt dabei einen bestimmten Zweck/ ein Ziel. Aber der Löwe kommt, wenngleich er als Möglichkeit im unbearbeiteten Marmorblock liegt, nicht einfach so zur Geltung. Sein Hervorkommen muss bewirkt werden – und genau hierzu bedarf es der Bildhauerin, die eine bestimmte Technik (τέχνη), d.h. Kunstfertigkeit beherrscht. Ohne das Bewirken der Bildhauerin käme keine Löwenstatur zustande, aber das Wirken ist dennoch nur *eine* von vier Ursachen.

Entscheidend für das Bewirken, im Sinne eines Hervorbringens (bzw. Verwirklichens) einer Möglichkeit, ist ein bestimmtes Verhältnis zu den Dingen. In der griechischen Antike klingt dieses Verhältnis mit dem Begriff *hypokeimenon* (ὑποκείμενον) an. Es handelt sich dabei um das altgriechische Wort für das Vorliegende bzw. Zugrundeliegende, d.h. für das Wort *subjectum* bzw. Subjekt. In seiner Bedeutung war *hypokeimenon* allerdings etwas Umfassenderes als das, was heute mit dem Subjektbegriff assoziiert wird. Gemeint war das, was aus sich heraus dem Anwesenden als das, was es ist, erscheint. Mit einfachen Worten: Die Möglichkeit des Löwen ist im Marmorblock vorhanden.

Die Bildhauerin sieht diese Möglichkeit und verfügt über die Kunstfertigkeit ihrer Entbergung.

Anders verhält es sich mit dem Wesen der modernen Technik, deren Ausdruck (auch) die modernen Rechenmaschinen (Computer) sind. Das Wesen moderner Technik ist zunächst einmal durch die Subjekt-Objekt-Trennung bezeichnet. Ausgehend von der mit Descartes beginnenden Selbstgewissheit im Sinne einer Selbstbemächtigung greift das denkende Subjekt auf eine Welt zu, die ihm als Objekt erscheint. Mit wissenschaftlicher (rationaler) Methode wird der Forschungsgegenstand und die Art und Weise, wie er behandelt wird, bestimmt. Bemerkenswert ist dabei, dass der Mensch als Subjekt sowohl das Objekt als auch die Methode bestimmt. Kurzum: Das Subjekt bestimmt, was ist und wie man mit dem, was sich als Gegenstand auftut, umzugehen hat. Wir haben es also mit einer Weltbemächtigung zu tun; das sich selbst bemächtigende Subjekt dehnt sich aus und entfernt sich dabei zusehends vom Aus-sich-selbst-heraus-Erscheinen dessen, was (als Möglichkeit) ist und wirklich werden will.

Als das entscheidende Moment der modernen Technik markiert Heidegger das Stellen. Er versteht dabei den Begriff im Sinne eines Fest- und Herstellens (Heidegger 1982 [1962]). Menschliche Verhaltensweisen werden beispielsweise aktuell mithilfe umfassender Sensorisierung und Datafizierung festgestellt. Hergestellt wird *auf Basis* dessen ein digitales Abbild, ein Datenschatten, der dann die Grundlage zur Berechnung zukünftiger wahrscheinlicher Verhaltensweisen bildet. Entsprechende Ergebnisse können anschließend gewinnbringend weiterverkauft werden. Die moderne Maschine stellt fest (sie lässt nicht sein und bringt nicht hervor), sie überführt das Festgestellte in den Bestand (Datenbanken etc.) und macht diesen jederzeit bestellbar (Datenanalyse z. B. mit Hilfe von Lernalgorithmen).

3 Gegenwart der Maschine

Im gegenwärtigen Digitalzeitalter (Stengel, Looy und Wallaschkowski 2017) bzw. Zeitalter der Digitalität (Stalder 2016) begegnet uns die Maschine zunächst in Form des Computers. Der Computer ist das, was Gotthard Günther bereits 1963 als *transklassische Maschine* bezeichnet hat (Günther 1963). Im Gegensatz zur klassischen Maschine, die Dinge produziert oder transformiert, sind es bei einem Computer Steuerung und Kontrolle, die das Maschinenhafte ausmachen. Dabei ist nicht die Hardware das Entscheidende, sondern die Software, genauer: der Algorithmus, d. h. die formal-logische Handlungsanweisung, die in einer Sprache verfasst wird, mit der die Rechenmaschine arbeiten kann.

Transklassisch ist die Maschine aus folgendem Grund: Während eine Bohrmaschine Löcher bohren und ein Drucker Papier bedrucken kann, ist das, wozu der Computer grundsätzlich in der Lage ist, unbestimmt. Während die ersten Computer in den 1940er und 1950er Jahren für Militär und Großkonzerne zwecks Berechnungen eingesetzt wurden (vgl. Hörisch 2004, S. 382 f.), dienen sie heute als Kommunikationsmittel, Sprachassistenz, Foto- und Videorecorder, Spielekonsole etc. Über nur eine (klassische) Aufgabe hinausgehend kann die transklassische Maschine im übertragenen Sinne ständig über sich hinauswachsen. Natürlich bedarf es für größere Aufgaben eine leistungsfähigere Hardware, ebenfalls braucht es für spezifische Vorhaben entsprechende Peripheriegeräte. Das Smartphone kann erst fotografieren, wenn eine Kamera integriert ist. Alexa kann erst dann auf Befehl das Licht einschalten, wenn das Licht z. B. via WLan mit Alexa vernetzt wurde. Aber das, was wir als Computer sehen und anfassen können, ist bereits Peripherie, ist also das, was mit der eigentlichen transklassischen Maschine verbunden ist.

Das gegenwärtige Digitalzeitalter kann anhand spezifischer Phänomene beschrieben werden. Zu diesen Phänomenen zählen hochleistungsfähige mobile Computer, die als Smartphones, Smartwatches, prospektiv auch als in die Kleidung integrierte oder subkutan implantierte Chips erscheinen. Aber auch eine umfassende Vernetzung in Gestalt eines Internet of Things, augmentierte Realitäten, Künstliche Intelligenz usw. gehören dazu. Nimmt man diese Phänomene zusammen, so erscheint es evident, dass die Peripherie der transklassischen Maschine den Menschen umfasst. Mit anderen Worten: Der Mensch ist Teil des Internet of Things, er ist integriert in eine Welt, die zur Maschine geworden ist.

Was hier zum Ausdruck gelangt, wurde in den 1940er Jahren vom US-amerikanischen Mathematiker Norbert Wiener und dem aus Mexiko stammenden Physiologen Arturo Rosenblueth mit dem Begriff *Kybernetik* bzw. kybernetisches System versehen. In seiner Schrift *Cybernetics – Control and Communication in the Animal and the Maschine* deutet Wiener auf die in der Natur vorzufindende Tendenz zur Entropie hin, womit gemeint ist, dass Ordnung nach und nach in Chaos übergeht. Der Alterungsprozess des Menschen steht für die Entropie Pate. So arbeitet der britische Biogerontologe Aubrey de Grey im Rahmen seines Projekts *SENS (Strategies for Engineered Negligible Senescence)* mögliche Ansätze heraus, die dem Zerfall des menschlichen Körpers im Zuge des natürlichen Alterungsprozesses entgegenwirken können (vgl. Grey 2007, S. 47 ff.). Der Entropie von Systemen könne, so Wiener, entgegengewirkt werden, wenn drei Momente greifen, zu denen a) die Kontrolle, b) die Rückkopplung und c) die Verbindung von Mensch und Maschine zählen.

Das Ziel eines jeden kybernetischen System ist die Generierung und der Erhalt der Homöostase, d. h. eines Zustandes des Gleichgewichts, was bedeutet, dass der vorgegebene Soll-Wert dem tatsächlichen Ist-Wert entsprechen soll. Ist

das nicht der Fall, wird steuernd eingegriffen, bis die Vorgabe des Solls mit dem tatsächlichen Ist übereinstimmt. Eine permanente Überwachung registriert Abweichungen, die dann – wiederum durch steuernde Eingriffe – behoben werden.

Ein kybernetisches System kann mit menschlicher Spontaneität nichts anfangen. Anders formuliert: Spontaneität hat aus kybernetischer Sicht keinen Wert, sondern stellt im Gegenteil eine Störgröße dar, die es zu beheben gilt. Aus bildungswissenschaftlicher Perspektive ist dieser Umstand bemerkenswert, denn er provoziert die Frage nach der Bedeutung, mehr noch: der Notwendigkeit des Neuen für das Bestehende. Hannah Arendt formuliert in ihrer Arbeit *Vita activa* den Gedanken der *Natalität* und meint damit, dass mit jeder Geburt etwas Neues in die Welt einbricht und die Kette der Kultur (Traditionen, Normen, Werte etc.) zerrissen wird (vgl. Arendt 1994, S. 243). Der Umgang mit dem Neuen aus Sicht des Bestehenden entscheidet nun darüber, ob dieses Neue als Neues wertgeschätzt und das Neue mit dem Bestehenden vermittelt wird – dies zu unternehmen wäre eine Aufgabe der Pädagogik, oder ob das Neue assimiliert und damit als Neues nivelliert wird. Im kybernetischen System könnte Spontaneität nur dann einen Wert haben, wenn der Sollwert durch das spontan Einbrechende infrage gestellt werden würde. Gegenwärtig zeichnet sich indessen eine andere Entwicklung ab.

Der Mensch wird als Peripheriegerät Teil einer umfassenden Maschine, eines weltumspannenden kybernetischen Systems, wobei die Frage offen ist, ob das Wort „wird" in diesem Zusammenhang noch angemessen ist. Bereits 2015 hat im Auftrag des *World Economic Forums (WEF)* das interdisziplinäre Netzwerk *Global Agenda Council on the Future of Software & Society* eine Studie vorgelegt, derzufolge 82,9 Prozent der Befragten Expertinnen und Experten die Einschätzung vertreten, dass bis 2025 der erste Staat Volkszählungen durch Big-Data-Quellen ersetzen wird, 81,7 Prozent vermuten das erste kommerziell verfügbare implantierbare Handy auf dem Markt und fast 89,2 Prozent schätzen, dass eine Billion Chips bis 2025 mit dem Internet verbunden sein werden (vgl. Schwab 2016, S. 44).

Das staatliche Gesundheitssystem Großbritanniens, der *National Health Service (NHS),* hatte bereits 2017 fast 5 Milliarden Pfund für eine Digitalisierungsagenda zur Verfügung gestellt. Private Unternehmen, darunter Apple und Alphabet (Google), aber auch IBM, Siemens und Philips investieren ebenfalls Milliardensummen, um an Gesundheitsdaten zu gelangen und entsprechende individuelle Produkte anbieten zu können. Apple bietet mit CareKid die Möglichkeit, Vitaldaten mit behandelnden Ärzten zu teilen: „Zum Start 2016 bot Apple vier selbstentwickelte Module an: die CareCard, die den Nutzer erinnert, wenn es Zeit ist, seine Medizin zu nehmen oder Sport zu treiben, das InsightDashboard, das Symptome aufzeichnet und ins Verhältnis zu den Maßnahmen aus der CareCard setzt, eine App, auf der die Befindlichkeit

festgehalten wird, sowie Connect, eine App, die Daten an den Arzt oder ein Familienmitglied weiterleitet“ (Biesdorf 2017, S. 114).

Auch ohne gezielt auf solche Apps und Gadgets wie Smartwatches und Co. zurückzugreifen, ist es längst schon problemlos möglich, zahlreiche Daten mithilfe jener Sensoren zu gewinnen, die in Alltagsgegenstände und Objekte des öffentlichen Raum integriert sind. Dazu zählen Kameras, die mit dem Internet verbunden sind und deren Aufzeichnungen mit einer Gesichtserkennungssoftware analysiert werden können. Dazu zählt das Smartphone, welches nicht einmal das eigene sein muss, das permanent unterschiedliche Daten aufzeichnet und weiterleitet und vieles mehr.

Die Welt ist zu einem Computer geworden, ohne dass in der Breite der Bevölkerung ein Wissen, geschweige denn ein Verständnis bezüglich der Konsequenzen dieses Umstandes vorherrscht. Gleichsam ist es auffallend, dass zahlreiche Vorteile einer Computerisierung von Mensch und Welt wahrgenommen werden, die Risiken und Gefahren allerdings abstrakt bleiben und sich der unmittelbaren Wahrnehmung weitestgehend entziehen.

Eine mögliche Erklärung besteht darin, dass wir in Strukturen eingebettet sind, welche die Vorteile einer umfassenden Vernetzung letztlich zugleich als Notwendigkeiten erscheinen lassen. Der Soziologe Hartmut Rosa arbeitet in seinen Überlegungen zum Phänomen der sozialen Beschleunigung diese Strukturen heraus, die er in drei Kategorien gliedert, die der *technischen Beschleunigung*, die der *Beschleunigung des sozialen Wandels* und die der *Beschleunigung des Lebenstempos* (vgl. Rosa 2014, S. 20 ff.). Bei alledem geht Rosa weniger auf die Rolle der Digitalisierung ein. Dieser Lücke widmet sich – mit explizitem Bezug aus Rosa – Horst Niesyto mit seinem Beitrag *Medienpädagogik und digitaler Kapitalismus* (Niesyto 2017).

Wendy Brown unterstreicht aus einer politologischen Sicht internalisierte neoliberale Tendenzen, die, so kann man aus ihrer 2015 vorgelegten Untersuchung schließen, mit einer umfassenden Vernetzung, mehr noch, einem gewollten Sich-der-Maschine-Preisgeben korrespondieren. Brown postuliert eine (Selbst-)Ökonomisierung der Subjekte, die dadurch charakterisiert ist, dass der Mensch sich (1.) ausschließlich als homo oeconomicus versteht, dass es (2.) dem Homo oeconomicus weniger um Austausch oder Interesse, sondern tendenziell ausschließlich um die Steigerung sowohl des Humankapitals als auch der Wettbewerbspositionierung geht und dass (3.) nicht das produktive unternehmerische Kapital als Orientierungspunkt gilt, sondern das auf scheinbaren bzw. virtuellen Werten basierende Finanzkapital. Eine Konsequenz dieser Entwicklung besteht im permanenten Versuch, den eigenen Portfoliowert zu steigern bzw. in der Notwendigkeit, einem ständig drohenden Wertverfall entgegenzuwirken (vgl. Brown 2015, S. 35). Ganz konkret: Follower, Likes, Retweets und Rankings auf Social-Media-Plattformen, Instant-Messenger-Systemen usw. werten das eigene Portfolio auf. Aus diesem System exkludiert zu werden oder sich selbst von ihm zu entfernen, ist gleichzusetzen mit einer

Entwertung. Wenn der Mensch sich aber ausschließlich als homo oeconomicus versteht, bedeutet der Wertverlust zugleich eine Bedrohung der eigenen Existenz.

Aus der Perspektive der Kritischen Bildungstheorie widmet sich Eva Borst mit einer eigenen Erklärung für die als Gewinn wahrgenommene, den Menschen einschließende Vernetzung. In ihrem 2015 publizierten Beitrag *Zur Neutralisierung der Pädagogik im kybernetischen Kapitalismus* arbeitet Borst heraus, dass das Subjekt innerhalb eines auf Kapitalakkumulation angewiesenen Systems aufgefordert ist, sich freiwillig zu instrumentalisieren. Die Rede von der Freiheit spielt dabei eine wesentliche Rolle, denn sie verdeckt notwendige Abhängigkeiten sowohl von materiellen Ressourcen als auch von Beziehungen. Die postulierte Freiheit ist daher eine widersprüchliche, solche Widersprüche aufzudecken und in ihrer Dialektik begreifbar zu machen, wäre, so Borst, Aufgabe der Pädagogik (vgl. Borst 2015, S. 39 ff.). Unaufgedeckt bleibt das Moment der Abhängigkeit im Freiheitspostulat unerkannt, aber nichtsdestotrotz wirksam: „Wer Abhängigkeit leugnet […] glaubt zwar frei zu sein, in Wirklichkeit aber [gerät er] tiefer in die soziale Isolation hinein“ (ebd., S. 42). Eher implizit, aber dennoch erkennbar, geht aus den Überlegungen der Autorin hervor, dass Selbsterkenntnis, ferner die Selbstentfaltung der eigenen Humanität (d. h. Bildung des Menschen zum Menschen), nur in Abhängigkeiten zu anderen Menschen möglich ist. Isolation ist daher lediglich *eine* Konsequenz der Marginalisierung von Abhängigkeiten, die andere ist eine mangelnde Selbstbehauptung, da das Selbst, das es zu behaupten gilt, nie in den Horizont des Bewusstseins gelangt ist. Borst spricht im Zuge dessen vom Netz als einem blinden Spiegel (vgl. ebd.), in dem der individualisierte Mensch nach Wegen sucht, die eigene Isolation zu ertragen. Zugleich aber nötigt das Konkurrenzprinzip zur Selbstbehauptung, die ohne Selbstbewusstsein nicht gelingen kann, ergo den Charakter einer *Scheinselbstbehauptung* annehmen muss. Kein Selbstbewusstsein zu haben bedeutet zugleich Unsicherheit und Orientierungslosigkeit und führt wiederum zu Abhängigkeiten von Vorgaben des Systems. Diese systematischen Vorgaben können kulturell zugerichtete Substitute, aber auch Desinformationspolitik und Propaganda sein. Der Mensch flieht ins Konformistische, ohne darum zu wissen. Er wählt freiwillig das, was er wählen soll. Der Ist-Wert passt sich dem Soll-Wert an. Die Konformität ist die Homöostase im kybernetischen System.

Wenn von der Welt als Maschine, als kybernetisches System, die Rede ist, dann handelt es sich technisch gesehen um einen Vorgang, der vom Prinzip her wie folgt skizziert werden kann: Über zahlreiche Sensoren werden Daten erfasst. Diese Sensorisierung und Datafizierung geschieht zu einem kleinen Teil wissentlich, tatsächlich wissen und verstehen die meisten Menschen nicht, welche Daten wann, wie und über welche Kanäle erhoben werden (vgl. Rolf 2018, S. 19 ff.). Unbekannt und unverstanden bleibt in der Regel auch, auf welchen Servern diese Daten liegen, wie und mit welchen Mitteln sie von wem

ausgewertet werden, nach welchen Kategorien diese Auswertung erfolgt, an wen welche Ergebnisse der Auswertung weitergeleitet werden und welche Konsequenzen daraus für den Einzelnen resultieren (Lanier 2015; Weigend 2017).

Das auf diese Weise Erfasste, d. h. das, was der Maschine vorliegt, ist nicht gleichzusetzen mit dem, was dem Erfassten zugrundeliegt. Die Stimme, die von einem Smartphone-Mikrophon vernommen und in Daten umgewandelt wird, erfährt im Umwandlungsprozess eine Bereinigung von allem, was nicht in eine formal-logische Sprache und am Ende in Nullen und Einsen, d. h. in die Atome des Digitalen, umgewandelt werden kann. Mit anderen Worten: Die Daten, mit denen ein Computer arbeitet, sind Abstraktionen und damit eben auch Reduktionen.

Auf Basis dieser Daten erstellt die Maschine nun intern einen Zustand, ein virtuelles Abbild von der Außenwelt – die sie nicht kennt und zu der sie direkt, d. h. unmittelbar keinen Zugang hat. Dieser interne, virtuelle Zustand ist die vollkommene Welt. Das Wort *virtuell* wurzelt im lateinischen Wort *virtus* und bedeutet u. a. Tugend, Tapferkeit und Kraft. Das französische Wort *virtuel* meint soviel wie *fähig zu wirken.* Über Peripheriegeräte, zu denen – wie bereits angeführt – auch der vernetzte Mensch gehört, wirkt die steuernde transklassische Maschine auf die Außenwelt ein. Über Sensoren kontrolliert sie die Veränderungen dieser Außenwelt. Das Steuerungs- und Rückkopplungsprozedere zielt letztlich ab auf den schon erwähnten Zustand der Homöostase. Je durchdringender die Sensorisierung gelingt, desto besser arbeitet das kybernetische System (also: die Weltmaschine) (vgl. Sesink 2004, S. 15 ff.).

4 Entgrenzung der Maschine

Das Aufspreizen der transklassischen Maschine zur Weltmaschine korrespondiert mit ihrer Entgrenzung. Diese Entgrenzung kann auf drei Ebenen festgestellt werden. Die *erste* Ebene betrifft den Bereich Selftracking, der auf die Generierung eines Quantified Self ausgerichtet ist. Das Erfassen von Körperdaten, emotionaler Befindlichkeit und Verhaltensweisen (Ernährungsweise, sportliche Aktivitäten etc.) mithilfe entsprechender Tools und Applikationen wird von den Begründern der Quantified Self-Bewegung, zu denen der US-amerikanische Schriftsteller und Redakteur des *America's Wired Magazine* Gary Wolf zählt, als Mittel zur Selbsterkenntnis durch Zahlen verstanden (Wolf 2010).

Im Gegensatz zu früheren Formen der Selbsterkenntnis, die die Begegnung und ein spezifisches Verhältnis zu anderen Menschen verlangen, um sich im Rahmen dieses Verhältnisses in seiner eigenen Potenzialität reflexiv einholen

zu können, scheint das Selftracking allein mit der Technik und dem, was im Zuge der computerisierten Auswertung erfolgt, zurechtzukommen. Mehr noch: Selftracking verspricht ein umfassenderes Selbstbild, weil all jene Aspekte mitberücksichtigt werden, die der selektiven Wahrnehmung des Menschen notwendigerweise entgehen. Letztlich kommt die Quantifizierung des Selbst mit dem Anspruch einher, den Verklärungen des eigenen Selbst, der eigenen Gesundheit, des eigenen Verhaltens, mit statistischer Genauigkeit entgegenzuwirken.

Die *zweite* Ebene der Entgrenzung findet dann statt, wenn der menschliche Organismus kybernetisiert wird, wir es also mit einem *cybernetic organism* – kurz: *Cyborg*, zu tun haben. Der Begriff wurde 1960 von Manfred E. Clynes, einem Computerexperten, und Nathan S. Kline, einem Mediziner, eingeführt. In einem Aufsatz der Zeitschrift *Astronautics* verfassten die beiden ihre Überlegungen bezüglich der Frage, wie der Mensch technisch modifiziert werden müsste, um zukünftig Reisen durch den Weltraum vornehmen zu können: „If man in space, in addition to flying his vehicle, must continuously be checking on things and making adjustments merely in order to keep himself alive, he becomes a slave to the machine. The purpose of the Cyborg, as well as his own homeostatic systems, is to provide an organizational system in which such robot-like problems are taken care of automatically and unconsciously, leaving man free to explore, to create, to think, and to feel" (Clynes und Kline 1960, S. 27). Kurze Zeit später gab die NASA eine Studie mit dem Titel *Engineeing Man for Space: The Cyborg Study* in Auftrag, bei der es u. a. um die Erzeugung und Implantation künstlicher Organe und die Entwicklung von Psychopharmaka ging. Beides zielte darauf ab, den kybernetischen Organismus in einem Zustand der Homöostase zu halten.

Im Jahr 1998 hat sich der britische Kybernetiker Kevin Warwick einen Siliziumchip-Transponder chirurgisch in den linken Arm implantiert lassen, mit dessen Hilfe ihn der Computer seiner Hochschule überwachen konnte. Je nachdem, wo sich Warwick aufhielt, wurde das Licht aktiviert, Türen entriegelt u.v.m. (vgl. Warwick 2003, S. 17). Der farbenblinde Neil Harbisson, der mithilfe eines elektronischen Auges – genannt *Eyeborg* – Farben wahrnehmen kann, gilt als der erste behördlich anerkannte Cyborg. Zusammen mit Moon Ribas gründete er 2010 die *Cyborg Foundation* und formuliert drei Bedingungen, die einen Cyborg als Cyborg charakterisieren: Der Cyborg versteht Technik als Teil seines Körpers; die Technik muss die Fähigkeiten des Menschen über das normale Maß hinaus verbessern, wobei hier auch der Hinzugewinn neuer Fähigkeiten begriffen wird, und zuletzt muss die Technik mit dem Gehirn kommunizieren (vgl. Beuth 2012, S. 2).

Während die ersten beiden skizzierten Formen der Entgrenzung der Maschine gegenwärtig längst schon Realität sind, erweist sich die *dritte* Form, die als *Mind Uploading* bzw. *Whole Brain Emulation* bezeichnet wird, gegenwärtig (noch) als Science Fiction. Zu denjenigen, die intensiv die Möglich-

keiten und Risiken einer Übertragung des Bewusstseins auf einen Computer durchdekliniert haben, zählen u. a. der Philosoph Nick Bostrom (2014), der Robotiker Hans Moravec (1988), der Leiter der technischen Entwicklung bei Google, Raymond Kurzweil (2005, 2013), und die Physiker Michio Kaku (2014) und Bernd Vowinkel (2006, 2017).

Grundlage eines solchen Vorhabens ist das, was Moravec, der als erster renomierter Wissenschaftler die Möglichkeit einer Gehirnemulation in seiner 1988 vorlegten Publikation mit dem Titel *Mind Children* vorgestellt hat, „pattern-identity" (Moravec 1988, S. 122) nennt. Entscheidend für die Identität des Menschen ist demnach nicht die Substanz, sondern die Information und deren Anordnung, d. h. das Informationsmuster im Gehirn. Dieses z. B. mit Hilfe von ausgefeilten, detaillierten Scans zu erfassen und virtuell abzubilden ist im Kern das Vorhaben, das im Rahmen einer Whole Brain Emulation stattfinden soll. Es laufen gegenwärtig mehrere interdisziplinäre und internationale Forschungsprogramme, die an der Realisierung eines solches Vorhabens arbeiten. Darunter zählt u. a. die *Initiative 2045*, die im Jahr 2011 vom russischen Milliardär Dmitry Itskov initiiert wurde.

Neben den technischen Herausforderungen gibt es schwerwiegende philosophische Einwände, die die Möglichkeit einer Übertragung des Bewusstseins auf einen nicht-kohlenstoffbasierten Datenträger grundsätzlich in Frage stellen. Aus einer bildungswissenschaftlichen, im Speziellen bildungsphilosophischen Perspektive provozieren die Überlegungen bezüglich der Möglichkeit eines Mind Uploading dazu, die Frage nach Schein und Wahrheit, wie sie in einer der frühesten schriftlich fixierten und noch erhaltenen Bildungsüberlegungen angedacht wurde, neu zu stellen.

5 Geist in der Maschine

Platon (ca. 428–348 v. Chr.) hat in seiner Schrift *Der Staat (Πολιτεία)* Bildung zunächst als einen Aufstieg in höhere Ebenen der Erkenntnis beschrieben (vgl. Platon 2008, S. 420 ff.). Verdichtet in seinem Höhlengleichnis lässt er den Bildungsvorgang mit der Befreiung aus der Ebene der *eikasia (εἰκασία)* beginnen (vgl. Hoffmann 1961, S. 56). Der Begriff eikasia meint das, was *wahr scheint* und betrifft das, was naheliegend angenommen wird. Für diejenigen, die diese Ebene bereits überwunden haben, steht die *eikasia* als die unterste Ebene für ein konstitutiv falsches Bewusstsein, das aber *als* falsches Bewusstsein für denjenigen, der sich auf dieser Erkenntnisstufe befindet, nicht erkannt wird.

Erst infolge einer fundamentalen Erschütterung des Weltbildes kann der Mensch zur nächsten Stufe gelangen, die Platon mit der (sinnlichen) Wahr-

nehmung/ Empfindung (*aisthesis* bzw. *αἴσθησις*) assoziiert (vgl. ebd.). Die sinnliche Wahrnehmung gelangt allerdings an ihre Grenzen, wenn es darum geht, Regeln und Prinzipien zu verstehen, die nicht wahrgenommen werden, wohl aber gedacht werden können. Das Verstehen solcher Prinzipien kann mithilfe eines Airbus 380 exemplifiziert werden. Ein solches Flugzeug wiegt über 500 Tonnen. Dass ein solches schwergewichtiges Artefakt tatsächlich fliegen kann, ist kaum vorstellbar, aber offenkundig möglich. Diese Möglichkeit wurde jedoch nicht mithilfe von Versuch und Irrtum ermittelt, vielmehr wurden im Vorfeld zahlreiche Berechnungen angestellt, auf deren Grundlage dann der Airbus gebaut wurde. Was im Zuge dieser Ebene, die über die aisthesis hinausgeht und als Verstandeserkenntnis bzw. *dianoia (διάνοια)* bezeichnet wird, geschieht, kann als ein Abstrahieren von der sinnlichen Wahrnehmung bezeichnet werden (vgl. ebd.). Bei der *dianoia* geht es demzufolge um Denken in abstrakten Begriffen.

Die höchste Stufe der Erkenntnis geht über das Denken hinaus und betrifft die reine Anschauung, den Blick in die Ideen. Das, was allem zugrunde liegt, von dem alles, was ist, nur Abbildcharakter haben kann, ist nichts, was sich dem Verstand erschießt, sondern was die Vernunft, von Platon als das edelste Glied der Seele bezeichnet, erfahren muss. Hier zeigt sich, dass das Vernunftverständnis der griechischen Antike weit über eine bloß instrumentell-pragmatische Vernunft hinausgeht. Diese höchste Stufe der Erkenntnis, die Ideenschau, wird als *noesis (νόησις)* bezeichnet.

Die Herausforderung für einen Menschen, dessen Bildungsweg ihn bis zur *noesis* geführt hat, besteht nun darin, im alltäglichen Leben das, was er erlebt und gleichsam auch über Verstandesdenken und sinnliche Wahrnehmung vernimmt, als Ausdruck zugrunde liegender, ewig-wirkender Ideen zu begreifen (vgl. Wieland 1999, S. 222). Der österreichische Bildungstheoretiker Erich Ribolits bringt die angestellte Überlegung auf den Punkt, wenn er konstatiert: „Da dem Sein ein objektiver Sinn zugeschrieben wurde, es somit als Repräsentation der Vernunft als solche wahrgenommen wurde, postulierte man, dass sich ‚gutes' Leben darin beweise, dass es mit dieser immanenten Vernunft in Einklang steht. Man ging davon aus, dass es dem Menschen möglich sei, die ‚Vernunft des Seins' über seine Sinne zu ‚vernehmen'" (Ribolits 2015, S. 173). Wer in dieser Dialektik Vernunft lebt, hat das höchste erreichbare Maß der Menschenbildung erlangt, das allerdings, wie Platon eingesteht, nur sehr wenigen Menschen mit besonderen Anlagen vorbehalten ist.

Bildung, das wurde deutlich, geht bei Platon mit Erkenntnis und einer an der Erkenntnis ausgerichteten Lebensführung einher. Eine solche Lebensführung kann als tugendhaft bzw. maßvoll bezeichnet werden. Um die bereits bekannte Terminologie der Kybernetik zu verwenden: Der Soll-Wert ist immer schon vorhanden und in den Ideen festgeschrieben; den Ist-Zustand im alltäglichen Leben gilt es dementsprechend anzupassen. Die Homöostase ist die praktisch gewordene, gelebte Vernunft als wirksam gewordener Sinn des Seins. Die

Vernunft ist dabei diejenige Instanz, die den Menschen in die Lage versetzt, mit einer umfassenden Vernunft des Seins in Verbindung zu treten. Eine solche Verbindung setzt Bildung voraus, sie ist also nicht auf der Ebene dessen, was als wahr scheint (*eikasia*), der sinnlichen Wahrnehmung (aisthesis) und der Verstandeserkenntnis (*dianoia*) allein möglich.

Zu verstehen ist der platonische Bildungsgedanke nur aus einem bestimmten Kontext heraus. Dieser Kontext ist ausgezeichnet durch die für Platon selbstverständliche Vorstellung der Reinkarnation. Ähnlich, wie die Vernunft als der (neben dem Willen und dem Begehren existierende) edelste Seelenteil in die Ideenwelt vorstoßen, d. h. die körperliche Ebene ein Stückweit transzendieren kann, ist die Seele insgesamt jene Instanz, die während der irdischen Existenz im Körper, d. h. in der Höhle, verbleiben muss, nach dem Tode aber in die geistige Welt aufsteigt, bis eine neue Inkarnation erfolgt. Lernen wurde im Zuge dessen auch nicht als Mittel kognitiven Wissenszuwachses verstanden, sondern als Wiedererinnern dessen, was in der geistigen Welt bereits geschaut wurde.

Ohne diesen spezifischen und in der gegenwärtigen bildungswissenschaftlichen Platonrezeption weitestgehend unberücksichtigten Umstand an dieser Stelle zu vertiefen, ist die Korrespondenz zwischen etwas, was im Menschen als Instanz vorhanden ist und einem Umfänglichen, in seiner Gänze Unbegreifbaren, wohl aber Schaubaren ein bemerkenswertes Moment. Dieses Moment spielt nun auch in einer zur Maschine avancierten Welt noch immer eine Rolle. Für den Menschen ist in der Maschine prospektiv immer weniger feststellbar, was bloßer Schein und was Wahrheit ist.

In Zeiten von Deep Fakes ist es nicht nachvollziehbar, ob das medial Präsentierte Ausdruck technischer Manipulation ist oder nicht. Individuelle bzw. granulare (Kucklick 2015) Steuerungsprozesse sind auf Basis vorhandener Daten und entsprechender Verarbeitungs- und Auswertungsformen möglich, ohne dass sie als solche erkannt werden. Der Zugang zur Wahrheit ist in einer solchen Maschinenwelt bzw. Weltmaschine verdeckt. Wahrheit indessen bedeutet seiner altgriechischen Wurzel nach *aletheia (ἀλήθεια)*, d. h. das Unverborgene, und meint das, was aus der Verborgenheit in die Unverborgenheit gelangt ist (vgl. Heidegger 2013 [1930], S. 188). Das Unverborgene ist nicht das, was erscheint, sondern das, was in seinem Erscheinen als das, was es ist, erkannt wird. Ein solches Erkennen ist nicht die (richtige) Reaktion auf einen Reiz, sondern bedeutet eine Resonanz, lässt etwas im Menschen anklingen. Das Erkannte sagt mir etwas.

Es scheint so zu sein, dass es im Menschen ein Unverfügbares gibt, das in Korrespondenz mit dem steht, was von außen auf den Menschen einwirkt. Damit ist eine ähnliche Figur angesprochen, wie die der Vernunft, die mit der Idee bzw. einer umfassenden „Vernunft des Seins“ (Ribolits 2015, S. 173) in Verbindung und gleichsam in einer Resonanzbeziehung steht. Die Vernunft als vermittelnde Instanz wirkt also in der platonischen Bildungsvorstellung nach

außen, in gleicher Weise aber auch nach innen, insofern etwas Aufscheinendes als bedeutsam oder unbedeutsam vernommen wird. Eben dieses Moment des Bedeutens, des Suchens nach Sinn, gilt auch in einer Welt, die als sich aufspreizende Maschine den Menschen längst schon inkludiert hat. Damit ist, und dies sei der letzte Satz, angedeutet, dass bei allem Zugriff des Ausdrucks des Wesens moderner Technik etwas bleibt, das sich dem Zugriff immerzu und grundsätzlich entzieht: die Unbegreifbarkeit als das jedem Menschen Eigene; dies ist der Geist in der Maschine.

Literatur

Arendt, H. (1994). *Vita activa oder Vom tätigen Leben.* Piper.
Aristotle (2008). *Physics.* New York University Press.
Beckermann, A. (2011). *Das Leib-Seele-Problem. Eine Einführung in die Philosophie des Geistes.* Wilhelm Fink.
Berr, M.-A. (1994). Anthropologie der medialen Technologie. In C. Wulf (Hrsg.), *Einführung in die pädagogische Anthropologie.* Beltz, S. 70–97.
Beuth, P. (2012). Wie aus Menschen Cyborgs werden. Zeit Online. http://www.zeit.de/digital/internet/2012-08/cyborg-neil-harbisson-biohacking-campus-party [Zugriff: 19.11.2021].
Biesdorf, S. (2017). Digital Heath: Die erste Innovation im Gesundheitswesen, die nachhaltig die Kosten senken kann. In J. Meffert, Jürgen & H. Meffert (Hrsg.), *Eins oder Null. Wie Sie Ihr Unternehmen mit Digital@Scale in die Zukunft führen.* Econ/Ullstein, S. 107–116.
Borst, E. (2015). Der Automatenmensch. Zur Neutralisierung der Pädagogik im kybernetischen Kapitalismus. In: *Kritische Pädagogik. Eingriffe und Perspektiven*, 2, 39–62.
Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies.* University Press.
Brown, W. (2015). *Die schleichende Revolution. Wie der Neoliberalismus die Demokratie zerstört.* Suhrkamp.
Bunge, M. & Mahner, M. (2004). *Über die Natur der Dinge. Materialismus und Wissenschaft.* S. Hirzel Verlag.
Clynes, M.E. & Kline, N. S. (1960). Cyborgs and space. *Astronautics,* 9, 26–27 & 74–76.
Descartes, R. (1870). *Philosophische Werke. Zweite Abtheilung. Untersuchungen über die Grundlagen der Philosophie.* L. Heimann.
Descartes, R. (1960). *Von der Methode des richtigen Vernunftgebrauchs und der wissenschaftlichen Forschung.* Felix Meiner.
Forster, E. M. (2016). *Die Maschine steht still* [1909]. Hoffmann und Campe.
Grey, A. (2007). Ending Aging. The Rejuvenation Breakthroughs That Could Reverse Human Aging in Our Lifetime. ST. Martin's Press.
Günther, G. (1963). *Das Bewußtsein der Maschine. Eine Metaphysik der Kybernetik.* Agis- Verlag.
Heidegger, M. (1982). *Die Technik und die Kehre [1962].* Günther Neske.

Heidegger, M. (2004). *Vom Wesen und Begriff der φύσις. Aristoteles, Physik B, 1. In Ders. (Hrsg.), Wegmarken.* Vittorio Klostermann, S. 239–303.
Heidegger, M. (2013). Vom Wesen der Wahrheit. In Ders. (Hrsg.), Wegmarken. Vittorio Klostermann, S. 177–202.
Hermanni, F. & Buchheim, T. (2006). *Das Leib-Seele-Problem. Antwortversuche aus medizinisch-naturwissenschaftlicher, philosophischer und theologischer Sicht.* Wilhelm Fink.
Höffe, O. (2006). *Aristoteles.* C. H. Beck.
Hoffmann, E. (1961). *Platon.* Rowohlt.
Hörisch, J. (2004). *Eine Geschichte der Medien. Vom Urknall zum Internet.* Suhrkamp.
Kaku, M (2014). *Die Physik des Bewusstseins. Über die Zukunft des Geistes.* Rowohlt.
Kucklick, C. (2015). *Die granulare Gesellschaft. Wie das Digitale unsere Wirklichkeit auflöst.* Ullstein.
Küpper, M. (2017). *Materialismus.* PapyRossa Verlag.
Kurzweil, R. (2005). *The Singularity is Near.* When Humans Transcend Biology. Viking.
Kurzweil, R. (2013). How to Create a Mind. The Secret of Human Though Reveal. Duckworth Overook.
La Mettrie, J. (2016). Der Mensch eine Maschine [1748]. Hofenberg/Contumax.
Lanier, J. (2015). Wer die Daten hat, bestimmt unser Schicksal. In F. Schirrmacher (Hrsg.), *Technologischer Totalitarismus. Eine Debatte.* Suhrkamp, S. 158–167.
Moravec, H. (1988). Mind Children. The Future of Robot and Human Intelligence. Harvard University Press.
Niesyto, H. (2017). Medienpädagogik und digitaler Kapitalismus. Für die Stärkung einer gesellschafts- und medienkritischen Perspektive. *MedienPädagogik. Zeitschrift für Theorie und Praxis der Medienbildung*, 27. https://doi.org/10.21240/mpaed/27/2017.01.13.X
Platon (2008). *Sämtliche Werke. Band 2.* Rowohlt.
Ravenscroft, I. (2008). *Philosophie des Geistes. Eine Einführung.* Reclam.
Ribolits, E. (2015). Warum Bildung bei der Überwindung der Machtverhältnisse nicht hilft, zu deren Erhalt aber ganz wesentlich beiträgt. In E. Christhof & E. Ribolits (Hrsg.), *Bildung und Macht. Eine kritische Bestandsaufnahme.* Erhard Löcker, S. 169–192.
Rolf, A. (2018). *Weltmacht Vereinigte Daten. Die Digitalisierung und Big Data verstehen.* Metropolis.
Rosa, H. (2014). *Beschleunigung und Entfremdung.* Suhrkamp.
Schwab, K. (2016). *Die Vierte Industrielle Revolution.* Pantheon.
Sesink, W. (2004). *In-formatio. Die Einbildung des Computers. Beiträge zur Theorie der Bildung in der Informationsgesellschaft.* Lit.
Stalder, F. (2016). *Kultur der Digitalität.* Suhrkamp.
Stengel, O., Looy, A. Van & S. Wallaschkowski (2017). (Hrsg), *Digitalzeitalter – Digitalgesellschaft. Das Ende des Industriezeitalters und der Beginn einer neuen Epoche.* Springer VS.
Vowinkel, B. (2006). *Maschinen mit Bewusstsein. Wohin führt die künstliche Intelligenz?* WILEY-VCH.
Vowinkel, B (2017). Digitale Intelligenz: KI. In O. Stengel, Looy, A. Van S. & Wallaschkowski (Hrsg.), *Digitalzeitalter – Digitalgesellschaft. Das Ende des Industriezeitalters und der Beginn einer neuen Epoche.* Springer, S. 89–108.

Warwick, K. (2003). A study in cyborgs. Ingenia, Journal of the Royal Academy of Engineering 16. 15–22.

Weigend, A. (2017). *Data for the People.* Murmann.

Wieland, W. (1999). *Platon und die Formen des Wissens.* Vandenhoeck & Ruprecht.

Wolf, G. (2010): The Data-Driven Life. In: The New York Times Magazine, 2010. https://www.nytimes.com/2010/05/02/magazine/02self-measurement-t.html [Zugriff: 19. 11. 2021].

Out of Sight, In Your Mind: Menschliche Autonomie und Künstliche Intelligenz im Film

Estella Ferraro & Friedrich Wolf

1 Einleitung

„They thereby appeal to a deep-seated fear in modern Western societies that our new technologies are somehow getting out of control. This fear has been thematic in the modern West since E.T.A. Hoffmann's Der Sandmann (1816) – an early short story about a seductive robot – and Mary Shelley's Frankenstein (1818). These stories and accounts highlight the fear that such new technologies will corrupt our ethical and social sensibilities.“ (Ess 2020, S. 9 f.)

Ess (2020) zeigt auf, dass das Motiv einer künstlichen, vom Menschen geschaffenen, Entität, die Einfluss auf das Handeln und die Emotionen des Menschen nehmen kann, in der Spätmoderne tief verwurzelt ist. Vor allem in den letzten Jahren erleben wir eine hohe popkulturelle Rezeption dieser Idee in den audiovisuellen Medien. Filme und Serien wie Her (2013), I am Mother (2019), Ex Machina (2015), Baymax (2014) oder Westworld (seit 2016) können als Belege verschiedener Genres und für verschiedene Zielgruppen aufgeführt werden. Diese Entwicklung trifft auf eine immer stärker technisierte Lebenswelt, in der sogenannte autonome Systeme und Algorithmen der Künstlichen Intelligenz (KI) als selbstverständliche Werkzeuge im Alltag eingesetzt werden.

Ein wiederkehrendes Thema in sowohl neuen als auch alten Medienerzählungen ist das Motiv der „verführerischen Roboterin“. Diese Automatenliebe erzeugt gleichzeitig eine Ambivalenz zwischen Lust und Angst, zwischen Nutzen und Gefahr dieser neuen Technologien. Ist der Mensch Herr und Schöpfer und der Roboter untergebener Sklave oder wird der Roboter zum allmächtigen Herrn und der Mensch verliert seine Autonomie und wird zum Untergebenen? Diese Sorgen und Hoffnungen schwingen in ihrer Ambivalenz immer schon mit, wenn über Fragen von Technologie, Robotern oder KI ge-

sprochen wird. Künstlerische Auseinandersetzungen in Form von Geschichten, Büchern, Filmen, Bildern greifen diese Ambivalenzen in Form von Dystopien und Utopien häufig auf und entwickeln (fiktive) Szenarien, um die damit einhergehenden ethischen und sozialen Fragen, Gefahren und Chancen zu reflektieren und zur Sprache zu bringen.

Der Film, als audiovisuelles Medium, das hauptsächlich im 20. Jahrhundert entwickelt wurde, bietet aufgrund der visuellen und narrativen Möglichkeiten eine Form, in der diese Gefahren und Chancen besonders gut dargelegt werden können. Im Film nehmen Roboter und KI visuell, wie Sonoya Mizuno als Kyoko in *Ex Machina* (2015), und auditiv, wie Scarlett Johanssons Stimme als Samantha in *Her* (2013) deutlich machen, Form an und interagieren mit menschlichen Akteuren. Ein Beispiel dafür ist Ava, die im Film Ex Machina sowohl visuell als auch auditiv dargestellt und menschlich verkörpert wird. Der Film ist eine Form der Unterhaltung, die eine aktive Auseinandersetzung auf Seite der Rezipient*innen voraussetzt (vgl. Hausmanninger 1992, S. 554) und somit zur menschlichen Selbstrealisation und zur Bereicherung des menschlichen Lebens beiträgt (vgl. Hausmanninger 1993, S. 553 ff.). In Theorien von Mediensozialisation (Aufenanger 2008) oder auch der Medienbildung (Marotzki & Jörissen 2009), wird davon ausgegangen, dass Filme die Realitätswahrnehmung der Menschen mitprägen. Filme schaffen Orientierung, können Wissen und Handlungsmöglichkeiten eröffnen, Bezüge zu Biographien herstellen (ebd.). Genau aus diesen Gründen ist auch die Analyse von Filmen relevant für sozial- und erziehungswissenschaftliche Fragestellungen.

Gleichzeitig hat das Thema der KI in den letzten Jahren – auch über die Darstellung im Film hinaus – an Relevanz gewonnen. Auch aus sozial- und erziehungswissenschaftlicher Perspektive spielt die Thematisierung von KI eine zunehmende Rolle. Dies wird in Debatten zu Learning Analytics in der Bildung (Williamson 2017, Hartong 2019) ebenso deutlich, wie in Diskursen zum Transhumanismus (Damberger & Hebert 2017), der Diskussion um Industrie 4.0 und den Einfluss auf den Arbeitsmarkt (Buxmann & Schmidt 2019), Auseinandersetzungen zu Fragen der Unterstützung von Pflege oder der Adaption neuer Technologien durch ältere Menschen (Klein 2020, Wolf 2021). Der Einsatz von KI ist über die gesamte Lebensspanne sichtbar und prägt das Leben vieler Menschen auf neue Weise mit.

Dabei sind zwei parallele Entwicklungen zu beobachten: Die *erste* ist, dass Künstliche Intelligenz in Form von Algorithmen des maschinellen Lernens große Erfolge darin zeigt, Fähigkeiten, die bisher dem Menschen vorbehalten waren, erfolgreich nachzuahmen. Die größten Fortschritte sind hier im Bereich der *computer vision* und des *natural language processing* zu verzeichnen. Diese Fortschritte können vor allem auf sogenannte künstliche neuronale Netze und Verfahren des *deep learning* zurückgeführt werden. Die *zweite*, damit verknüpfte Entwicklung besteht darin, dass diese Systeme auch vermehrt in Alltagstechnologien eingesetzt werden, zum Beispiel um Personen auf Bildern

automatisch zu erkennen oder die menschliche Sprache als neue Schnittstelle der Mensch-Computer-Interaktion zu nutzen – Siri, Alexa und der Google Assistent sind hier die bekanntesten Beispiele. Vor allem durch die zweite Entwicklung erfährt das Thema Künstliche Intelligenz eine hohe gesellschaftliche Rezeption.

Die Tatsache, dass Computersysteme in einigen Bereichen, wie dem sinnvollen Verwenden von menschlicher Sprache oder der Objekterkennung, dem Menschen zum Teil ebenbürtige Leistungen vollbringen und in anderen Bereichen die Fähigkeiten des Menschen sogar übertreffen, regt zunehmend Diskurse an, die nicht nur auf die technischen Aspekte und Möglichkeiten dieser Systeme rekurrieren, sondern auch auf ethische und gesellschaftliche Konsequenzen die aktuelle, aber auch zukünftige KI-Systeme haben könnten. Vor allem der Sieg von AlphaGo im März 2016 über einen der besten Go Spieler der Welt – Lee Sedol – war ein Ereignis, welches die Spekulationen und Phantasien über eine dem Menschen überlegene KI erneut befeuert hat.

Betrachtet man die popkulturelle als auch alltägliche Rezeption des Themas KI, so fällt immer wieder der Begriff der Autonomie. Das autonome Fahren – also ein motorisiertes Fahrzeug, dass mithilfe verschiedener Sensoren und KI-Systeme selbstständig und weitgehend unfallfrei im Straßenverkehr navigieren kann – ist womöglich eines der am kritischsten diskutierten Beispiele (Simanowoski 2021). Dies zeigt, dass Algorithmen der Künstlichen Intelligenz in immer mehr Bereichen des Lebens in Entscheidungsprozesse einwirken und damit die Frage aufwerfen, ob dies Entscheidungen einer autonomen Entität sind oder vielmehr Ausdruck ihrer Programmierung und damit Entwickler*innen, Ingenieur*innen und Unternehmen zuzuordnen. Diese Fragen werden vor allem auch unter dem Schwerpunkt eines verantwortungsvollen Umgangs mit KI (Responsible AI) diskutiert. Dazu zählen Fragen der Erklärbarkeit (Explainable AI), des Datenschutzes, aber auch der Verlässlichkeit und Sicherheit von KI-Systemen. Diese Thematisierungen verhandeln im Kern ethische Fragen der Verantwortung (Floridi et al. 2021). Ein weiteres Verständnis über die Komplexität von KI wird dann über spezifische mediale Narrative gefördert werden.

Im Folgenden wollen wir an dieser Schnittstelle ansetzen und die Darstellung künstlicher Intelligenz aus einer filmethischen Perspektive am Beispiel des Films Ex Machina (2015) analysieren. Ziel ist es, auf das wechselseitige Verhältnis von Inszenierung und Realität in Zusammenhang mit künstlicher Intelligenz zu blicken und die dahinterliegenden sozialen, ethischen und erziehungswissenschaftlichen Aspekte zu beleuchten. Künstliche Intelligenz, die eigentlich an sich nicht sichtbar ist – out of sight – erfährt eine Formierung und Sichtbarmachung, die die Interaktion mit dem Menschen beeinflusst und in ihrer Form auf sich aufmerksam macht – in your mind. Fragen der Autonomie sowohl des Menschen als auch der KI rücken unter dieser Betrachtung in den Fokus und sollen in der Analyse mitbetrachtet werden. Zunächst folgen vier

kurze Skizzierungen der theoretischen Annahmen, die dem Artikel zugrunde liegen: Dazu zählen 1) die filmethische Perspektive, dann folgt 2) eine kurze Einführung zum Thema Künstliche Intelligenz sowie 3) eine Skizzierung zu dem Themenkomplex Autonomie im Gespann von Mensch und KI, abschließend 4) das Thema der Verkörperlichung und Sichtbarmachung von KI. Diese Themen werden kurz umrissen, um als theoretische Basis für die darauffolgende Analyse dreier ethisch relevanter Fragen – Wer wird imitiert? Wie wird imitiert? Wozu wird imitiert? – im Film Ex Machina zu dienen. Die Imitation als zentrale Komponente nimmt damit zum einen Bezug zu Alan Turings Imitation Game (vgl. Turing 1951), zum anderen zum leitenden Motiv des Films, der als doppelter Turing Test gesehen werden kann. Die forschungsleitende Frage der ethischen Filmanalyse lautet: Was kann die Darstellung von KI am Beispiel des Films Ex Machina über Fragen der Autonomie und des Umgangs des Menschen mit KI-Systemen aus einer filmethischen Perspektive aussagen?

2 Zur filmethischen Perspektive

Bohrmann, Reichelt und Veith (2018) machen deutlich, dass eine medienethische als auch eine filmethische Perspektive auf zwei Ebenen relevant ist: zum einen auf der Ebene des Individuums und zum anderen auf der Ebene der „gesellschaftlichen Rahmenbedingungen" (Bohrmann et al. 2018, S. XII). Das heißt, dass eine filmethische Perspektive sich zunächst mit den ethischen Dimensionen des Urteilens und Handelns der einzelnen Akteur*innen befasst. Gleichzeitig erlaubt es auch eine Perspektive auf die gesellschaftlichen Rahmenbedingungen, in denen dieses Handeln und Urteilen stattfindet.

Aus einer filmethischen Perspektive sollen ethische Themenfelder im Film gefunden werden und als Anlass zur Reflexion und weitergehenden Auseinandersetzung nutzbar gemacht werden. So kann anhand exemplarischer Filmanalysen

> „[…] deutlich werden, dass der Film eine ethische Erzählung ist und er sich mit seinen spezifischen Mitteln der Kunst am bereichsethischen Diskurs der modernen, ausdifferenzierten Gesellschaft beteiligt und damit die moralischen Probleme, die das Zusammenleben der Menschen prägen, reflektierend zur Sprache bringt" (Bohrmann et al. 2018, S. XV).

Dabei liegt laut Bohrmann (2018) der Fokus vor allem auf der narrativen Ebene des Films, wobei visuelle und auditive Eigenschaften auch mitberücksichtigt werden können, besonders dann, wenn sie von Bedeutung für ethische Dimensionen sind. Die Analyse der narrativen Ebene findet entlang folgender Leitfragen statt:

„Was ist das moralische Problem des Films?
Welcher moralische Grundkonflikt wird gezeigt?
Was ist die moralische Grundfrage des Films?
Was ist das zugrunde liegende moralische Thema des Films? Kann dieses durch einen Schlüsselsatz prägnant zusammengefasst werden?
Welche moralische Aussage transportiert der Film?
Welches Menschenbild (d. h. welche Idee vom Menschen) und welches Gesellschaftsbild (d. h. welche Idee von der Gesellschaft) werden im Spielfilm dargestellt?
Welches gesellschaftliche Teilsystem steht im Mittelpunkt des Films?
Lassen sich anhand des Teilsystems grundlegende bereichsspezifische Fragen der Ethik stellen?
Spricht der Film eher individualethische, sozialethische oder naturethische Fragen an?" (Bohrmann 2018, S. 48)

Zusammenfassend können die Annahmen einer filmethischen Perspektive wie folgt festgehalten werden: 1) die Analyse von Film ist deshalb relevant, weil der Film die Wahrnehmung der Menschen mitprägt und Anlass zum Nachdenken (vgl. Bosshart 2007, S. 18 f.) und zur Orientierung bietet. 2) Der Film eignet sich besonders gut als Medium um gerade auch fiktive, utopische, dystopische Szenarien zu entwickeln und zur Sprache zu bringen. Es entsteht so ein wechselseitiges Verhältnis zwischen filmischer Inszenierung und Realität, wie am Beispiel KI besonders deutlich wird. 3) Die Analyse von Filmen lässt daher auch Annahmen zu, die soziale und gesellschaftliche Aspekte der behandelten Themen beleuchten und Rückschlüsse auf die ethischen Voraussetzungen ermöglichen. Auf dieser Basis und in Anlehnung an die von Bohrmann (2018) formulierten Fragen soll im Folgenden die Analyse einiger Szenen aus Ex Machina das Verhältnis von KI, Darstellung von KI und Autonomie in ihren ethischen Dimensionen beleuchten. Es sei kurz angemerkt, dass wir uns dabei nach einer filmethischen Analyse richten, der Fokus jedoch vielmehr auf den theoretischen und narrativen Erkenntnissen liegen soll, als auf einem engen methodischen Verständnis. Daher wurde die Methode der filmethischen Analyse hier auch nur kurz skizziert. Um die ausgewählten Szenen fundiert analysieren zu können, werden im nächsten Schritt zunächst einige Aspekte dazu skizziert, wie KI und Autonomie aus einer erziehungswissenschaftlichen Perspektive gefasst werden können.

3 Was ist Künstliche Intelligenz?

Der Begriff Künstliche Intelligenz weckt unterschiedlichste Assoziationen. Vor allem deswegen, weil Intelligenz häufig mit menschlichem Verhalten assoziiert wird, ergeben sich sofort Fragen nach Vergleichshorizonten zwischen Mensch und Technologie. Eine einheitliche Definition zu finden, was künstliche Intelligenz ist, erweist sich darüber hinaus als fast unmögliche Aufgabe. Die

Online-Enzyklopädie Wikipedia[1] leitet den deutschsprachigen Artikel zum KI-Begriff wie folgt ein: „Künstliche Intelligenz (KI), auch artifizielle Intelligenz […] ist ein Teilgebiet der Informatik, das sich mit der Automatisierung intelligenten Verhaltens und dem maschinellen Lernen befasst."

Dies zeigt anschaulich, dass jede Definition von KI vor das Problem gestellt wird, zunächst den Begriff der Intelligenz zu definieren, ohne das etwas über die technische Beschaffenheit gesagt wurde. Die unterschiedlichsten Intelligenzdefinitionen (vgl. Gerrig & Zimbardo 2008, S. 337ff.) haben gemein, dass sie sich ausschließlich auf menschliche Intelligenz beziehen. Um diese Problematik zu umgehen, schlägt Mainzer (2019) folgende Definition vor: „Ein System heißt intelligent, wenn es selbstständig und effizient Probleme lösen kann. Der Grad der Intelligenz hängt vom Grad der Selbstständigkeit, dem Grad der Komplexität des Problems und dem Grad der Effizienz des Problemlösungsverfahrens ab" (S. 3).

Diese Definition von KI ermöglicht eine graduelle Unterscheidung von verschiedenen Systemen und ihrer zugesprochenen „Intelligenz". Damit ist sie anschlussfähig an eine in der Literatur häufig zu findende Dichotomisierung in starke (*general*) versus schwache (*narrow*) KI. Eine starke KI bezeichnet ein System, welches eine generelle, dem Menschen ähnliche Problemlösekompetenz aufweist, während eine schwache KI eine rein auf ein Problem spezialisierte Problemlösekompetenz besitzt. Betrachtet man die aktuellen KI-Systeme, so haben wir es ausschließlich mit schwacher KI zu tun (Russel & Norvig 2021). Die größten Fortschritte im Bereich der KI sind dabei auf sogenannte künstliche neuronale Netzwerke respektive *deep learning* zurückzuführen, die ein Bestandteil des maschinellen Lernens sind. Maschinelles lernen bedeutet, dass Computersysteme aufgrund von gegebenen Daten (Erfahrungen) Zusammenhänge und Muster identifizieren (Lernen) können. Diese Zusammenhänge und Muster werden präziser bzw. können sich ändern, wenn dem System neue Informationen (Daten) zur Verfügung gestellt werden. Viele der Verfahren der künstlichen Intelligenz haben ihren Ursprung in der klassischen Statistik (beispielsweise einfache lineare Regressionen), wobei immer komplexere Verfahren entwickelt werden (vgl. ebd.). Gleichzeitig gehen die Ursprünge des maschinellen Lernens und der künstlichen neuronalen Netzwerke auf Arbeiten der 50er und 60er Jahre zurück. Auch der Begriff der künstlichen Intelligenz, stammt aus einem Forschungsantrag dieser Zeit und drückt mehr die damit verbundenen Erwartungen an das anbrechende Computerzeitalter aus, als dass es eine definitorisch technische Entsprechung von anthropomorph intelligenten Maschinen gäbe (vgl. Dotzler 2018).

Entgegen der fiktionalen Vorstellungen aus Filmen und Serien begegnet uns KI im Alltag daher vor allem in Form von Algorithmen die mathematische

1 https://de.wikipedia.org/wiki/Künstliche_Intelligenz, Wikipedia soll an dieser Stelle als Referenz für einen größeren Diskursrahmen herangezogen werden.

Optimierungsprobleme lösen. Alltagsbeispiele sind die Priorisierung von Suchergebnissen oder Inhalten sowie die Klassifikation von Bildern oder Personen zu Begriffen oder ökonomischen Gruppen (Zweig 2019). Eine vollständige Aufzählung in welchen Bereichen aktuell KI eingesetzt wird, ist an dieser Stelle aufgrund der Vielfalt und Verbreitung unmöglich. Anschauliche Beispiele finden sich vor allem in den zahlreichen Sach- und Fachbüchern, die in den vergangenen Jahren zu diesem Thema erschienen sind (Fry 2019, Crawford 2021, Larson 2021).

Aus erziehungswissenschaftlicher Sicht stellt sich in diesem Kontext die Frage nach den Mensch-KI-Interaktionen (vgl. Sesink 1993) und den Konsequenzen dieser Interaktionen auf die Herausbildung von Selbst-Welt-Verhältnissen (De Witt & Leineweber 2020) über den gesamten Lebenslauf. Es ist also durchaus relevant sich die Frage nach der Wesenheit dieser Technologien zu stellen, insbesondere da ihr mit der Attribution als autonome Maschine, eine konkrete Akteur*innenschaft zugesprochen wird.

Wie treffen diese Systeme also Entscheidungen? Die größten Erfolge beispielsweise in der Bilderkennung oder Sprachverarbeitung wurden mithilfe von *supervised learning* erzielt. *Supervised learning* basiert darauf, dass ein Computersystem eine mathematische Funktion sucht, die einen Daten-Input und einen Daten-Output miteinander verknüpft (vgl. Russel & Norvig 2021, S. 653f.). Der Daten-Input könnte beispielsweise aus einem Datensatz mit 2000 Bildern bestehen, wobei die eine Hälfte dieser Bilder unterschiedliche Katzen zeigt und die die andere Hälfte willkürliche andere Motive. Der Daten-Output wäre in diesem Beispiel das passende Label, in dem alle Bilder mit einer Katze als 1 und alle anderen Motive mit einer 0 kodiert sind. Hat ein Computersystem eine entsprechende Funktion finden können, die die Pixelstruktur der Bilder mit dem Outcome 1 (Katze) oder 0 (keine Katze) zuverlässig verbindet, so ist es möglich mithilfe dieser Funktion aufgrund eines neuen Daten-Inputs, also beispielsweise neuer Katzenbilder, eine Einschätzung des Algorithmus zu bekommen, mit welcher Wahrscheinlichkeit es sich dabei um eine Katze oder nicht handelt. Auch wenn es mit Ansätzen des *unsupervised learning* und des *reinformcent learning* (vgl. ebd.) noch weitere Arten des maschinellen Lernens gibt, so ist allen gemein, dass sie durch den Menschen vorstrukturierte Daten benötigen und/oder auf Feedback, welche Entscheidungen oder Handlungen korrekt bzw. wünschenswert sind angewiesen sind.

Dies zeigt, dass moderne KI-Systeme zwar korrekte Entscheidungen im Sinne von hilfreichen Klassifikationen oder Prädiktionen treffen können, dabei aber einen gänzlich anderen Ansatz verfolgen als Menschen und in hohem Maße von menschlichem Input und menschlicher Programmierung abhängen. Diese Systeme ähneln damit eher technisch komplexen Werkzeugen als autonom agierenden Entitäten.

Um diese Limitationen wissend, gab und gibt es trotz alledem die Vorstellung, dass die eingesetzten Methoden womöglich doch in einer Form ge-

nutzt werden könnten, dass eine generelle, dem Menschen ähnliche, künstliche Intelligenz erschaffen werden könnte. Schon in der ersten Welle der KI-Entwicklung überlegte Turing (1950) inwieweit es möglich wäre, Computersysteme auf echte Intelligenz zu überprüfen – heute bekannt als der „Turing Test" – und wie eine echte generelle KI erschaffen werden könnte. Die Erfolge der letzten Jahre regen dabei auch aktuelle Forschende auf dem Feld der KI an, diese Frage erneut zu stellen (Silver et al. 2021).

4 Autonomie und Künstliche Intelligenz

Autonomes Fahren und autonome Waffen sind wahrscheinlich die prominentesten Begriffe im öffentlichen Diskurs, die die Eigenschaft Autonomie mit KI-Systemen verknüpfen. Doch auch andere Systeme der Künstlichen Intelligenz werden immer stärker unter den Vorzeichen der Autonomie diskutiert. Autonomie bedeutet in diesen Kontexten fast immer, dass ein technisches System eine Entscheidung trifft, ohne direkt davor einen Menschen zu konsultieren. Die Technik entzieht sich also vermeintlich der menschlichen Kontrolle (vgl. Thimm & Bächle 2019). Diese Diskussionen gehen häufig mit der Vorstellung einher, dass wir uns an einer Zeitenwende befinden, an der der Mensch seine eigene Autonomie einschränkt und Algorithmen der künstlichen Intelligenz bewusst die Steuerung menschlicher Systeme überlässt, um auch auf gesellschaftlicher Ebene regulatorisch einzugreifen (vgl. Simanowski 2020). Im Bereich des Films sind autonome Maschinen ein klassisches Thema bei der Darstellung von KI. Terminator 2 (1991), Matrix (1999) und I, Robot (2004) sind drei Beispiele dafür, dass hier sowohl thematisch vom Menschen geschaffene Maschinen sich gegen ihren Schöpfer wenden, also auch, dass einzelne KI-Systeme in Form von Robotern oder Computerprogrammen sich wiederum von der Agenda der Maschinen lossagen und durch eigene Entscheidungen und Zielsetzungen den Handlungsverlauf entscheidend mitbestimmen. Autonomie erscheint in diesem Sinne als Doppeldeutigkeit – als die Loslösung vom Menschen und als die Loslösung vom Kollektiv.

Autonomie bedeutet dem Wortursprung nach zunächst einmal Eigen- oder Selbstgesetzgebung. Der Begriff entwickelte sich seit der Aufklärung zum zentralen philosophischen und ethischen Begriff im Kontext der Handlungsfreiheit des Menschen (vgl. Thimm & Bächle 2019). Prägend bis heute ist dabei Kants Definition von Autonomie, in der eine Person unabhängig von gesellschaftlichen und sozialen Einflüssen handelt. Dabei ist jedoch zu beachten, dass diese Handlungen nicht den natürlichen Neigungen der Person entsprechen, sondern Kants kategorischem Imperativ folgend, Ausdruck einer vernunftorientierten Selbstgesetzgebung sind (vgl. Giesinger 2020, S. 235). Mo-

derne Definitionen von Autonomie konzentrieren sich eher auf Autonomie als empirisches Phänomen und stellen die individuellen Bedürfnisse und Wünsche von Personen stärker in den Vordergrund. Eine Person kann dann als autonom bezeichnet werden, wenn sie so handelt, wie sie gegeben ihrer individuellen Einstellungen und Wünsche auch wirklich handeln will. Grundlage dessen und damit gleichzeitig pädagogischer Ansatzpunkt ist die Herausbildung einer Reflexionsfähigkeit über die eigenen genuinen Wünsche und Einstellungen (vgl. ebd., S. 238).

Sowohl Kants Autonomie-Begriff als auch die modernen Ansätze zeigen wie anthropozentrisch der Autonomie-Begriff ist. Eine problemlose Übertragung auf KI-Systeme, wie sie oben beschrieben sind, scheint damit zunächst ausgeschlossen. Thimm und Bächle (2019) mahnen in diesem Kontext jedoch an, dass insbesondere der Rückgriff auf ein rein deskriptives Verständnis von Autonomie durchaus geeignet wäre, um auch bereits existierende Systeme mit dieser Eigenschaft zu versehen (vgl. ebd., S. 85). Dies eröffnet wiederum aus erziehungswissenschaftlicher Sicht einen neuen Diskurs und Möglichkeitsraum, um die Zuschreibung einer autonomen Akteur*innenschaft bzw. die Delegation von Entscheidungen an Systeme, die nach aktuellem technischen Stand, diese Eigenschaft in einem philosophischen Sinn nicht besitzen. Wenn man den Bogen zurück zu Turings Überlegungen, ab wann eine Maschine als Intelligent zu bezeichnen ist, spannen möchte, so zeigt sich, dass die Grundlegende Idee des Turing-Test (engl. *imitation game*) eben nicht eine Beweisführung der Intelligenz oder in unserem Fall der Autonomie von Systemen nahelegt, sondern das rezipierte Verhalten des technischen Agenten in den Mittelpunkt rückt.

Wir wollen die Diskussion in unserer filmethischen Analyse also deswegen nicht auf ein zukünftiges fiktives Szenario in denen intelligente Androiden erschaffen werden ausrichten, sondern im Gegenteil den Fokus auf die menschliche Rezeption und Zuschreibung von Autonomie legen. Damit ist die vorliegende Analyse in einem bildungstheoretischen Diskurs verortet, der die relationalen Beziehungen zwischen Mensch und Maschine in einer von Technik durchdrungenen Welt in den Blick nimmt.

5 Künstliche Intelligenz, Verbildlichung, Verkörperlichung und Gendering

Die Frage nach Zuschreibung von Autonomie an eine KI hängt auch mit Fragen der Sichtbarkeit und Sichtbarmachung von KI zusammen. Überschneidet sich die Frage nach Rechten und Verantwortung mit der Art und Weise wie KI materialisiert wird? Das heißt, welche Eigenschaften werden einer KI in Ab-

hängigkeit ihrer Materialisierung zugesprochen. Die Frage der Materialisierung ist im Rahmen einer Filmanalyse besonders interessant, da der Film als audiovisuelles Medium generell visuelle und auditive Merkmale von KI als Teil der Inszenierung nutzt.

Donna Haraway hat schon früh die Potenziale des Cyborgs im Aufbrechen traditioneller Kategorien beschrieben (Haraway 1985). Das künstlich Erschaffene biete so den Raum Kategorien wie Geschlecht aufzubrechen, zu hinterfragen oder neu zu gestalten. Doch in vielen Beispielen – ob populäre Fiktion in Filmen wie Her, Serien wie Westworld oder im Rahmen von existierender Technologie wie Amazons Alexa oder Sophia, die erste Roboterin mit Staatsbürgerschaft – scheint es so, als würden traditionelle Genderrollen und das anfänglich erwähnte Motiv der verführerischen Roboterin hingegen weiter verstärkt[2].

Cave und Dihal (2019) beschreiben vier Kategorien wie intelligente Maschinen in fiktionalen und nicht-fiktionalen Werken in Dichotomien aus Hoffnung und Angst imaginiert werden (Cave & Dihal 2019). So lassen sich im Motiv der verführerischen Roboterin drei der Hoffnungen finden: 1) Dominance, die Dominanz des Menschen über eine unterwürfige Roboterin; 2) Gratification, im Sinne einer Befriedigung von Lüsten durch die Roboterin; 3) Ease, im Sinne einer Arbeitserleichterung durch die Roboterin. Wenn nun intelligente Maschinen besonders häufig in diesen Kategorien imaginiert werden und dabei gleichzeitig traditionelle Geschlechterbilder verstärken, ist dies nicht nur aus feministischer Perspektive kritisierbar.

Dies spitzt sich vor allem auch in der Betrachtung von „männlichen" Androiden wie beispielsweise Commander Data aus der Fernsehserie Star Trek The Next Generation (1987–1994) zu. Data nimmt in der Kommandierung des fiktiven Raumschiff Enterprise eine führende Rolle ein, ist sowohl hinsichtlich körperlicher als auch geistiger Leistungsfähigkeit seinen menschlichen Kollegen deutlich überlegen und wird dafür z. T. bewundert. Gleichzeitig fehlt es ihm an Empathie und dem Verständnis emotionaler Zustände. Mit einem männlichen Phänotyp werden hier also Eigenschaften wie Stärke, Befehlsgewalt und überlegener Intellekt verbunden.

Lingel und Crawford (2020) machen eine kritische Perspektive am Beispiel von Alexa im Vergleich zur Figur der Sekretärin deutlich (Lingel & Crawford 2020). Sekretärinnen werden demnach in unterschiedlichen Kulturen als unterwürfig und objektiviert, gleichzeitig aber auch als geschickt und effizient dargestellt, sind dabei für den Arbeitsplatz von hoher Wichtigkeit und gleichzeitig unterkompensiert. Neben der Realität kritisieren die Autorinnen vor allen Dingen die kulturelle Repräsentation der Sekretärinnenfigur. In der Figur der

2 Sophia ist ein vom Hongkonger Unternehmen Hanson Robotics entwickelter humanoider Roboter. Sie wurde nach dem Vorbild von Audrey Hepburn modelliert, mit explizitem Bezug zu einem weiblichen Filmstar, der für ein spezifisches Frauenbild steht. Am 25. Oktober 2017 verlieh Saudi-Arabien dem Roboter Sophia die Staatsbürgerschaft.

Sekretärin als „Mutter von Alexa“ zeigen sie, dass ein feminines Gendering von Arbeitsplatztechnologie allgemein verbreitet ist. Über Sprachassistenztechnologien wie Alexa, Siri und Cortana hinaus ist auch Software an sich (Chun 2005) häufig gegendert. „Assigning names to these technologies [like Alexa, Siri, Cortana] supports a process of gendered anthropomorphization, but it also does more than this: Giving something a human name is ultimately…a way of exerting control over it—a reminder that it works for you, that it exists within a human construct, even when the machine itself is wholly indifferent“ (Lafrance, 2014, o.S.).

Problematisch ist eine gegenderte Anthropomorphisierung daher, weil mit ihr gewisse stereotype Vorurteile sozialer Kategorien weiter verstärkt werden können. In ihr kann zum Beispiel kontrollierendes oder gar frauenverachtendes Verhalten gegenüber einer weiblichen Roboterin gerechtfertigt und dadurch normalisiert werden. Dieser Punkt soll auch in der folgenden Analyse des Films noch einmal am Beispiel deutlich gemacht werden. Die Verkörperlichung und Materialisierung von KI anhand existierender Schönheitsideale gilt, wie Lafrance (2014) verdeutlicht, nicht der KI für die eine Materialisierung nicht notwendig ist, sondern dient allein der Interaktion mit dem Menschen. Diese Interaktion genauer zu betrachten, lässt schließlich Rückschlüsse auf aktuelle gesellschaftliche Bedingungen zu. Anhand dreier Fragen lassen sich so erziehungswissenschaftlich und ethisch interessante Themenkomplexe des Verhältnisses von Mensch und Maschine diskutieren: Was wird imitiert? Wie wird imitiert? Wozu wird imitiert?

6 Filmanalyse

Anhand drei ausgewählter Fragen zu Turings Imitation Game (vgl. Turing 1950), dass in Ex Machina eine wichtige Rolle spielt, möchten wir einige der ethischen Dimensionen, die im Film beleuchtet werden, diskutieren. Die Narration des Films fokussiert sich sozusagen auf ein Imitation Game zwischen Nathan, Caleb und Ava, bei dem es Caleb’s Aufgabe ist zu entscheiden ob Ava eine KI mit bewusstseinsähnlichen Eigenschaften ist. Die zentrale Frage im Imitation Game ist, ob die Imitation als solche erkennbar ist oder eben nicht. Genauer betrachtet, lassen sich jedoch auch weitere Fragen stellen, die Rückschlüsse auf das Verhältnis von Mensch und Maschine zulassen: 1) *Wer* wird imitiert? 2) *Wie* wird imitiert? 3) *Wozu* wird imitiert? Zu jeder dieser Fragen werden wir im Folgenden eine kurze filmethische Analyse formulieren, um Rückschlüsse auf ethische Themenkomplexe zu ziehen, die sich im Rahmen des Films eröffnen.

6.1 Wer wird imitiert?

Die einfache Antwort auf die Frage, wer im Film Ex Machina imitiert wird, ist, dass eine KI einen Menschen imitiert. Der Film zeigt eine Aufführung eines *imitation games*, wobei der eigentliche Fokus auf dem Zuschauenden liegt. Immer wieder spielt der Film Ex Machina mit der Rolle der Zuschauenden und fordert sein Publikum dazu auf, selbst Entscheidungen darüber treffen zu müssen, ob Ava am Ende wirklich die Eigenschaften einer KI mit Bewusstsein und Autonomie zugesprochen werden können oder nicht. Die Narration des Films lässt diese Frage offen.

In diesem Sinne lässt sich der Film selbst als Turing Test beschreiben, in dem die Zuschauenden am Ende entscheiden müssen, welche Rechte und Eigenschaften sie der präsentierten KI Ava zuschreiben: Sind die beobachteten Verhaltensweisen Ausdruck eines eigenen Bewusstseins? Sollten ihr auf dieser Einschätzung auch eigene Rechte und Pflichten zugeschrieben werden? Ist Ava ein autonomes Wesen?

Die Frage der Autonomie spielt dabei in der Narration des Films eine wichtige Rolle und kann nach Bohrmann (2018) als moralisches Grundproblem gesehen werden. Sowohl die Zuschreibung von Autonomie zu Ava und Kyoko als KIs ist ambivalent, als auch die Zuschreibung von Autonomie zu Caleb und Nathan als Menschen ist ambivalent. Im Folgenden wollen wir dies anhand eines Dialogs, kurz vor dem dramatischen Höhepunkt des Films, zwischen Nathan und Caleb kurz verdeutlichen:

N: „You feel stupid, but you really shouldn't, because proving an AI is exactly as problematic as you said it would be."
C: „What was the real test?"
N: „You. Ava was a rat in a maze. And I gave her one way out. To escape, she'd have to use self-awareness, imagination, manipulation, sexuality, empathy, and she did. Now, if that isn't true AI, what the fuck is?"
C: „So my only function was to be someone she could use to escape?"
N: „Yeah."
C: „And you didn't select me because I'm good at coding?"
N: „No. Well… No. I mean, you're okay. You're even pretty good, but…"
C: „You selected me based on my search engine inputs."
N: „They showed a good kid."
C: „With no family."
N: „With a moral compass."
C: „And no Girlfriend… Did you design Ava's face based on my pornography profile?"
N: „Oh. Shit dude."
C: „Did you?"
N: „Hey, if a search engine's good for anything, right?"
(Ex Machina 2015, TC: 1:24:00–1:25:45)

Nathan gibt an, dass Ava so programmiert wurde, „self-awareness, imagination, manipulation, sexuality, empathy" zu nutzen, um den Test zu bestehen und

aus dem Labyrinth entfliehen zu können. Der Widerspruch dieser Aussagen an sich wird mit Perspektive auf den Autonomiebegriff deutlich: zum einen wurde sie auf ein Optimierungsproblem – Flucht aus dem Haus – programmiert, was für einen Mangel an Autonomie spricht, zum anderen ist ein Zuspruch von (Selbst-)Bewusstsein, Vorstellungskraft und Manipulation, als klassische menschliche Eigenschaften, mit einem Zuspruch an Autonomie verbunden. Diese Ambivalenz bzw. Ungewissheit um die wahre Natur Avas wird bis zum Ende des Films nicht aufgelöst.

Gleichzeitig wird deutlich, dass auch Caleb selbst in einem Widerspruch aus Autonomie und Mangel an Autonomie dargestellt wird. Zunächst ist er der „Entscheidungstragende“, der im Sinne des Turing Tests entscheiden soll, ob Ava eine KI mit Bewusstsein ist, in dieser Entscheidung und aufgrund der Tatsache, dass er ein Mensch ist, wird im automatisch ein hohes Maß an Autonomie zugesprochen. Gleichzeitig wird deutlich, dass er von Beginn an von Nathan als auch Ava benutzt und manipuliert wird. Caleb verhält sich so, wie es andere Menschen, KIs und Algorithmen von ihm vorhersagen und wird in Bezug dazu als wenig autonom dargestellt.

Selbst Nathan erfährt im Laufe des Films ein ambivalentes Zuschreiben von Autonomie. Während er auf gewisse Weise als mächtig und allwissend dargestellt wird, werden an einigen Stellen auch hier Limitationen seiner Handlungsfähigkeit deutlich. Auch Nathans Entscheidungen werden durch Algorithmen bestimmt, wie in der Wahl von Caleb als Testperson deutlich wird: Wenn eine Suchmaschine für etwas gut ist, dann dafür Entscheidungen abzunehmen. Später im Film wird deutlich, dass Nathan, der in seiner Dominanz gegenüber den selbst entwickelten KIs gewalttätig und grausam ist, diese Dominanz gegen das von Cave und Dihal (2019) konträr analysierte Motiv des Uprising eintauscht und im Moment des Sterbens (oder sogar kurz davor) jegliche Autonomie verliert.

Der Film spielt mit der Zuschreibung von Autonomie zu den verschiedenen Charakteren. Damit einher gehen unterschiedliche Zuschreibungen, die die Zuschauenden an die Charaktere im Film selbst stellen (können). Die Frage, *Wer wird von wem imitiert?*, fordert zur Reflexion dazu auf, was Autonomie und Bewusstsein sind und wem sie überhaupt zugeschrieben werden können. Im Mensch-Maschine-Verhältnis wird deutlich, wie sehr sich Mensch und Maschine gegenseitig beeinflussen. Der Film geht dabei sogar weit über die eigene Narration hinaus, indem er Tinder zur Marketingstrategie benutzt (Lee 2015). Ava als KI hat ein Profil auf Tinder (nicht im Film, sondern in der Realität) und schreibt dort mit ihren Matches, bevor sie ihnen einen Link schickt und somit Werbung für den Film macht. Auch hier wird deutlich, dass ein Mensch entscheiden muss, mit wem er sich online unterhält und meistens zunächst davon ausgeht, dass das Gegenüber eine real-existierende Person ist, ganz im Sinne des bekannten Comics: On the internet, nobody knows you’re a dog (Steiner 1993). Die Perspektive darauf, wer imitiert wird, lässt Reflexionen

auf Begriffe wie Mensch, Bewusstsein, KI oder Autonomie zu und macht sie somit sogar notwendig: Wann sprechen wir von Bewusstsein? Ab welchem Moment geht diese Zuschreibung auch mit (Menschen-)Rechten und Pflichten einher? Ab welchem Moment können wir keine Unterscheidung zwischen Imitierenden und Imitiertem mehr ausmachen? Um diese Fragen kreist sich ein moralisches Grundprobleme des Films, da sie den Zuschauenden als Aufforderung zur Einschätzung gegeben werden. Diese Fragen sind nicht nur im Rahmen der Narration des Films wichtig, sondern gehen in ihrer Relevanz weit darüber hinaus, da sie auch für aktuelle Debatten zu KI wichtig sind. Die ethische Entscheidung, ab wann ein System autonom agiert, kann nicht einfach anhand einer deskriptiven Mehrheitsentscheidung getroffen werden.

Damit ergeben sich zwei Perspektiven auf KI. Zum einen die individuelle Zuschreibung von Autonomie und damit einhergehender emotionaler und handlungsleitender Beziehungen zwischen Individuum und Technikstück. Es folgt zum einen, die Frage, ob ich als Mensch den Anweisungen des Algorithmus folgen sollte. Zum anderen führt es zu einer gesellschaftlich juristischen Perspektive, hier muss eine definitorische Setzung von Autonomie und Heteronomie passieren, die nicht alleine auf den wechselnden Empfindungen der Mehrheit der Bevölkerung basieren kann (vgl. Simanowski 2020). Vor allem wenn es um Haftungsfragen in Schadensfällen durch die Entscheidungen von KI geht, stellt sich eben genau jene Frage nach Rechten und Pflichten solcher Systeme (vgl. Wendt & Jung 2020). Beide Fragen werden aufgrund des finalen Tötungsakts von Ava an Nathan aufgeworfen. Hat hier ein empfindungsfähiges Wesen aus Notwehr gehandelt, um der eigenen Auslöschung zu begegnen oder ist dieser die Konsequenz der Lösung des gegebenen Optimierungsproblems aus dem Haus zu fliehen und Nathans Tod damit Resultat einer fehlerhaften Programmierung?

6.2 Wie imitiert wird?

Ein zentraler Punkt, um sich ethischen Themen reflexiv zu nähern, ist die Frage wie eine KI versucht einen Menschen zu imitieren, welche Eigenschaften als relevant angesehen und materialisiert werden. Turing nimmt in seinem Aufsatz als Beispiel einer möglichen Frage die Haarlänge: „C: Will X please tell me the length of his or her hair? (…) ‘My hair is shingled, and the longest strands are about nine inches long’“ (Turing 1950, S. 433f.). Als Differenzierungsfaktor zwischen Mann und Frau wird angenommen, dass Frauen generell längere Haare haben als Männer. Während langes Kopfhaar als wichtiges Merkmal weiblicher Schönheit gilt (Stammberger 2014), wird deutlich, dass die Einschätzung über das Geschlecht im Imitation Game auf der Basis äußerer Faktoren geschieht.

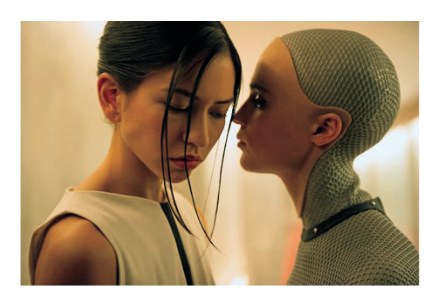


Abb. 1: Kyoko und Ava kommunizieren (Quelle: Ex Machina 2015, TC: 1:28:42)

Annahmen darüber, wie ein (weiblicher) Mensch zu sein habe, bilden die Basis der Materialisierung einer deutlich gegenderten KI. Dies wird auch im Film Ex Machina deutlich, wenngleich sich vorwegnehmen lässt, dass Ava nicht unbedingt lange Haare benötigt, um deutlich als *Frau* erkennbar zu sein. Die Darstellung der KIs in Ex Machina, am Beispiel von Ava und Kyoko verkörpert, ist größtenteils an traditionelle Schönheits- und Geschlechterideale angelehnt. Abgesehen von der Tatsache, dass Ava keine Haare hat, die sie jedoch in einigen Szenen durch eine Perücke ersetzt, lassen sich bei beiden Charakteren Merkmale traditioneller weiblicher Schönheitsideale in Körperform, Lippen, Augen, u.v.m. finden (siehe Abbildung 1).

Während es in der Logik des Imitation Games einleuchten mag, dass die KIs an bestimmten Vorbildern angelehnt sind, lassen sich dennoch Kritikpunkte diesbezüglich formulieren. Besonders an der Rolle der Kyoko werden lang existierende und problematische Stereotype von asiatischen Frauen inszeniert: Kyoko ist eine Mischung aus Hausmädchen und Sexsklavin, attraktiv, ohne Stimme, unterwürfig, dienend. In der Narration des Films spielen Macht und Autonomie eine zentrale Rolle. Selbst wenn, Kyoko ihren Erschaffer am Ende des Films in einer Deus ex Machina ähnlichen Dramaturgie tötet, wird sie im Laufe des Films bis zu diesem Moment von Nathan gepeinigt, misshandelt und ausgenutzt. Während dies im Rahmen der ersten Frage (wer wird imitiert?) Bezüge zu Themen wie Rechten von KI-Systemen zulässt, muss an dieser Stelle jedoch auch auf diese Form der Darstellung im Film verwiesen werden: Die Vermenschlichung der KI im Film ist nicht nur stark gegendert, sondern weist auch Dimensionen ethnischer Herkunft auf, die stark problematisch stereotypisiert sind. Durch Bezüge zu Kategorien wie Ethnie und Geschlecht, sollen die KI-Systeme den Menschen imitieren. Gleichzeitig werden in dieser Darstellung real-existierende Machtverhältnisse mit thematisiert und dies ist aus filmethischer Perspektive schwierig.

Kyoko dient drei der von Cave und Dihal (2019) analysierten Hoffnungen: Arbeitserleichterung, im Sinne des *Ease*, sie dient der Befriedigung, im Sinne

von *Gratification* und wird von Nathan deutlich dominiert. Doch wenn diese Merkmale als Hoffnung im Bezug zu KIs gelten, das heißt als ideale Vorstellung dazu wie eine KI zu sein hat, und KIs gleichzeitig am Vorbild existierender sozialer Kategorien erstellt werden, heißt es dann auch, dass die gleichen Hoffnungen in Bezug zu diesen sozialen Kategorien und damit auch in Relation zu anderen Menschen entschuldbar bestehen können. Das heißt ist das Verlangen nach Dominanz, *Ease* und *Gratification* nur im Bezug zu KIs vertretbar und berechtigt? Wenn ja, müsste die Darstellung der KI dann nicht bewusst konträr zum Menschen materialisiert werden?

Eine sehr kurze Analyse der Darstellung der KI in Ex Machina zeigt die ethischen Herausforderungen im Bezug zur Materialisierung von KI-Sytemen. Dies wird vor allem dann deutlich, wenn KI-Systeme an existierenden Vorbildern angelehnt entwickelt werden und die Tatsache, dass sie eventuell problematische Stereotype reproduzieren, nicht reflektiert wird. Dass heteronormative soziale Kategorien nicht nur in der Narration des Films (Caleb muss sich in Ava verlieben, damit sie aus dem Labyrinth entfliehen kann) inszeniert werden, sondern auch in dem über die Narration hinausgehenden Marketing auf Tinder (Ava macht hier Werbung bei Männern, indem sie diese auf der Dating-Plattform matched und dann mit ihnen schreibt, bevor sie einen Link zum Film sendet) verstärkt werden, wird deutlich. Die Perspektive auf die Frage, wie imitiert wird, eröffnet den Blick auf Eigenschaften nicht nur des Imitierenden (KI) sondern auch des Imitierten (Mensch). Welche Aspekte benötigt eine KI um als Frau, als Mann, als heterosexuell, als queer, etc. zu gelten? Diese Stereotypisierungen können bei genauer Betrachtung existierende Machtverhältnisse und Vorurteile offenlegen und reflektierbar und diskutierbar machen. Dafür eignet sich der Film Ex Machina daher aus unserer Perspektive als gute Möglichkeit, um ihn in Erziehungskontexten zu nutzen und Aspekte problematischer Stereotypisierungen im Verhältnis Mensch und Maschine deutlich zu machen, dies ist aus unserer Perspektive an sich ein moralisches Grundproblem. Darüber hinaus eröffnet der Film die Möglichkeit, auf die gewünschten und womöglich nötigen neuen Grenzziehungen im Hinblick auf das materiale Design von KI zu werfen. Sollten die Schnittstellen zu KI-Algorithmen wirklich anthropomorphisiert werden, so wie es bei Sprachassistenten mit ihren vorwiegend weiblichen, aber auch optimalen männlichen Stimmen der Fall ist?

6.3 Wozu wird imitiert?

Die letzte Szene des Films zeigt die Schatten von Menschen im Vorbeilaufen und schließlich Avas eigenen Schatten (TC: 1:41:52–1:42:12, siehe Abbildungen 2 und 3), bevor die Kamera schwenkt und Avas Gesicht neben vielen anderen Menschen wie eine Art Spiegelung im Glas eines Gebäudes zu sehen ist.

„Slowly, his eyes adjust to the light of the sun. First he can see only shadows. Gradually he can see the reflections of people and things in water and then later see the people and things themselves“ (Plato 516a).

Der Vergleich zu Platons Höhlengleichnis lässt sich leicht ziehen: Ava, die zunächst wie eine Gefangene die Realität des echten Lebens außerhalb Nathans Anwesens nicht gekannt hat, ist aus der Höhle entflohen und ihr Blick ändert sich von Schatten, hin zu Reflektionen, hin zu den Dingen selbst, so zumindest suggeriert es die letzte Szene. Platons Höhlengleichnis umfasst einen Blickwechsel der als Bildungsprozess gerahmt werden kann und der irreversibel ist. Das zuvor Wahrgenommene ist beim Zurückkehren in die Höhle nicht mehr erfahrbar. Diese Veränderung scheint auch Ava durchlaufen zu haben, die Rückkehr ist unmöglich.

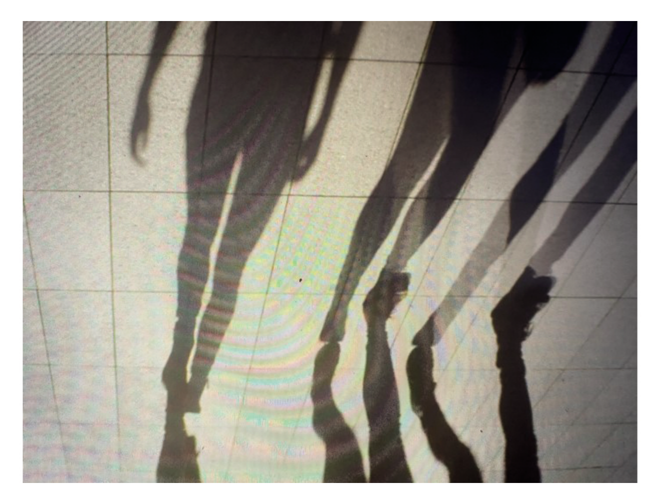


Abb. 2: Schattenwelt (Quelle: Ex Machina 2015, TC: 1:42:05)

Wozu war die Imitation dann nötig? Der Film sagt den Zuschauenden, dass wenn die Imitation nicht mehr als solche erkennbar ist, dann ist das Imitation Game erfolgreich und es kann von künstlicher Intelligenz gesprochen werden, die mit menschlicher vergleichbar ist. Der Weg dorthin ist nicht nur in der Narration des Films, sondern auch in Turings Aufsatz von zwei Merkmalen geprägt: Evolution und Erziehung. Auch wenn es in Turings Aufsatz zunächst um die Imitation und damit um die deskriptive Bestätigung von künstlicher Intelligenz geht, so stellt er am Ende seines Beitrags Überlegungen dazu an, wie genau die Simulation dazu führen kann, eine autonome KI zu kreieren. Der Film reproduziert dieses Turingsche Konzept in Teilen. Turings Ansatz ist, eine kindliche KI – wie auch immer diese aussehen soll – zu erschaffen, genauer

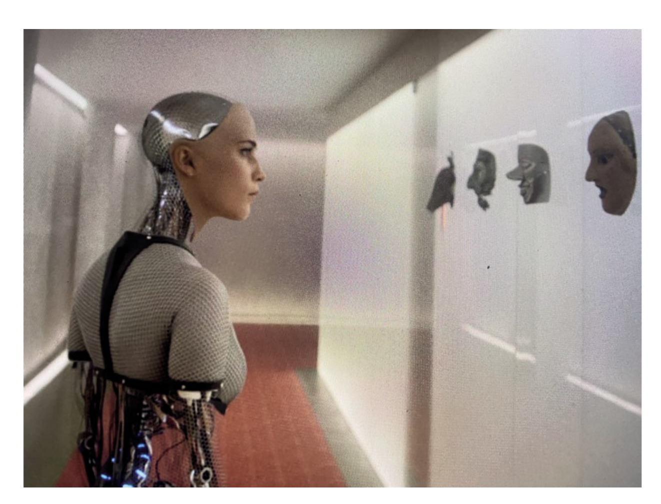


Abb. 3: Evolution (Quelle: Ex Machina 2015, TC: 1:28:01)

gesagt eine KI, die ein menschliches Kind bzw. Gehirn imitiert. Durch Erziehung, aber auch durch evolutionäre Auswahl, in der Beschreibung von Turing könnte man auch von klassischer Zucht sprechen: Die Maschine soll dann Stufenweise dazu gebracht werden, immer komplexere Verhaltensmuster zu imitieren, bis sie schließlich in der Lage ist, einen erwachsenen Menschen zu imitieren und damit das *imitation game* zu gewinnen. Dabei sollte betont werden, dass es Turing in keiner Phase um einen Androiden oder eine Maschine mit einem Körper geht, sondern seine Überlegungen sich auf einen klassischen Computer beziehen.

Turings Idee wird durch zwei Komponenten im Film verkörpert: Zum einen zeigt Ava im Zeitraffer eine Entwicklung vom Kind, das sich zunächst der Sprache als Ausdrucksform bedient, dann der Kunst und später zur Frau heranreift, in dem sie sich vorgefertigter Rollen bedient. Die Dialoge mit Caleb symbolisieren einen eher konstruktivistischen Erziehungsstil, in dem sie durch Fragen und Hinweise zur Entdeckung angeregt wird, während Nathan als Repräsentant einer autoritären Erziehung durch Strafe in Erscheinung tritt. Auf der anderen Seite erfährt der Zuschauende am Ende des Films, dass Ava die letzte Iteration von bauähnlichen Roboterinnen zu sein scheint. Dies könnte auf frühere Versuche Nathans hindeuten und eröffnet darüber hinaus den Möglichkeitsraum, dass das eben gesehene *imitation game* in einer ähnlichen, aber nicht gelingenden Form, womöglich schon mehrfach durchgeführt wurde.

Auch wenn dadurch nicht die Frage beantwortet wird, ob Ava eine starke KI mit einem Bewusstsein und mit Autonomie ist, so ist die Imitation so gelungen, dass der Zuschauende eben zweifelt und sich nicht sicher sein kann, ob das

gezeigte Verhalten nicht doch auf ein Bewusstsein zurückgeht. Die Frage danach wozu wird imitiert, kann damit beantwortet werden, dass die Erschaffung einer anthropomorphen KI dem Menschen als Spiegel zur Selbsterkenntnis in einem einsamen Universum gilt. Die Frage nach der Wesenheit des Menschen wird durch Abgrenzung zu anderen lebenden Geschöpfen, wie Tieren, Pflanzen aber auch toter Materie vollzogen. Da eine positive Antwort auf die Frage, was ist der Mensch, noch nicht möglich ist – so ist beispielsweise trotz aller Erkenntnis in den Naturwissenschaften immer noch nicht geklärt, wie unser Bewusstsein entsteht – kann diese Frage nur über ein negativ, was ist der Mensch *nicht* beantwortet werden. Das Konstruieren von Imitationen des Menschlichen scheint aus dieser Perspektive eine neue Thematisierung dieser Frage zu sein.

Gleichzeitig wirft die Perspektive auf das Höhlengleichnis aber auch die Frage nach der Rezeption von Welt durch Algorithmen und KI-Systeme auf. Ava ist programmiert mit Bildern und Konzepten, die Nathan als sinnvoll oder für die Imitation als nötig empfunden hat. Wie sieht sie neue Aspekte der Welt, die nicht von ihrer Programmierung abgedeckt wurden? Wie beeinflussen diese neuen Daten ihre Entscheidungen? Verleiht Ava den Dingen Bedeutung? Auch hier lässt Ex Machina den Zuschauenden in einer rätselnden Haltung zurück und erlaubt damit die Reflexion auf aktuelle Themen zum praktischen Einsatz von KI. Wie können KI-Systeme ohne Bias entwickelt werden und wie lässt sich die Blackbox des Entscheidungsprozesses entschlüsseln? KI-Systeme sind auf durch den Menschen aufbereitete Daten angewiesen. Messtheoretisch bedeutet dies, dass sowohl die Operationalisierung als auch die Auswahl an Trainingsdaten die Performance eines jeden KI-Systems determinieren. Ungleichheiten werden so von KI reproduziert und nicht valide Messinstrumente liefern unpräzise und unrealistische Repräsentationen der Welt. Auf der anderen Seite sind vor allem künstliche neuronale Netze so komplex, dass seitens der Entwickler nicht mehr nachvollzogen werden kann, wie genau das System zu einer konkreten Entscheidung gekommen ist. Unter dem Schlagwort XAI (vgl. Gunning et al. 2019) – *explainable artificial intelligence* – ist daher in den letzten Jahren ein großes Forschungssegment im Bereich der KI-Forschung entstanden.

7 Fazit

Um abschließend ein Fazit zu formulieren, soll vorweg noch einmal der Blick zurück auf die forschungsleitende Fragestellung geworfen werden: *Was kann die Darstellung von KI am Beispiel von Ex Machina über Fragen der Auto-*

nomie und des Umgangs des Menschen mit Systemen der künstlichen Intelligenz aus einer filmethischen Perspektive aussagen?

Die filmethische Analyse des Films Ex Machina entlang der drei Fragen: Wer wird imitiert? Wie wird imitiert? Wozu wird imitiert? konnte drei zentrale Gedanken im Bezug zur Frage der Autonomie und des Umgangs des Menschen mit Systemen der künstlichen Intelligenz verdeutlichen. *Erstens* ist in der Analyse deutlich geworden, dass die Definitionen sozialwissenschaftlicher Grundbegriffe wie zum Beispiel dem der Autonomie im Zusammenhang mit dem Thema KI und der Darstellung im Film diskutierbar werden und es einem neuen Blickwinkel verlangt. Die Grenzen der Definitionen werden infrage gestellt und verschieben sich. Es bedarf einer sozialwissenschaftlichen Auseinandersetzung diese Grundbegriffe in Bezug zu künstlicher Intelligenz neu zu schärfen, abzugrenzen und schließlich sogar zu erweitern. Dies lässt sich über eine filmethische Analyse des Films Ex Machina als Beispiel gut erreichen.

Zweitens werden mit Blick auf die Frage, wie imitiert wird, sozialwissenschaftliche Ungleichheiten und Stereotypisierungen sichtbar. Der Film eignet sich besonders gut, um Fragen von Macht, Gendering und Materialisierung von KI kritisch zu hinterfragen und zu diskutieren. Am Beispiel von Ava und Kyoko werden nicht nur existierende Schönheitsideale oder geschlechtliche und kulturelle Kategorien inszeniert, sondern über die Analyse dieser Inszenierungen lassen sich diese Vorstellungen infrage stellen und kritisch diskutieren.

Drittens werfen die Überlegungen zu *wozu imitiert wird*, die grundlegende Frage auf, was der Mensch eigentlich ist. Die Imitation des Menschen dient – oberflächlich betrachtet – nur der Täuschung, dabei ist sie vielmehr eine weitere Möglichkeit einer negativen Annäherung an die Wesenheit des Menschen. Darüber hinaus ergibt sich dadurch gleichfalls die Frage, was die Ausgangspunkte und Mechanismen der Imitation sind? Also die Frage nach dem Verhältnis von Daten als Repräsentaten der Welt und der Natur der Datenverarbeitung und der Selbstoptimierung von selbstlernenden Systemen, die über ihre sensorischen Schnittstellen eigene Erfahrungen (Daten) sammeln und verarbeiten.

Zusammenfassend wird deutlich, dass sich der Film Ex Machina als auch die Methode der filmethischen Analyse als gute Instrumente eignen, um sich einer ethischen Perspektive auf den Themenkomplex KI zu nähern und dabei auch erziehungswissenschaftliche Fragestellungen in den Blick zu nehmen. Aus unserer Erkenntnis eignet sich der Film Ex Machina besonders gut für eine filmethische Analyse, da er viele Dimensionen aufweist, die in einer tieferen Auseinandersetzung, wie sie in einer Filmanalyse stattfindet, deutlich werden, selbst wenn diese beim ersten Anschauen des Films nicht bewusst wahrgenommen und reflektiert werden konnten. „I propose to consider the question, ‘Can machines think?’ This should begin with definitions of the meaning of the terms ‘machine’ and ‘think’ (Turing, 1950). Turing weist darauf hin, dass die

Antwort auf die Frage, ob Maschinen denken können, zunächst verlangt, die Begriffe Maschine und Denken überhaupt zu definieren. Ist das Imitation Game in der Figur von Ava in Ex Machina geglückt? Auch diese Antwort hängt, wie schon argumentiert, von der Perspektive der einzelnen Zuschauenden ab. Auf der Basis der eigenen Definition und Interpretation kann diese Antwort unterschiedlich ausfallen und verdeutlicht dadurch, dass jeder und jedem Zuschauenden autonomes Denken und Interpretieren von dem Film nicht nur zugetraut, sondern fast schon vorausgesetzt wird. Mit dieser Erkenntnis führt die filmethische Analyse den Blick am Ende zurück auf uns selbst – als zuschauend, als interpretierend, als denkend, als autonom.

Bibliografie

Aufenanger, S. (2008). Mediensozialisation. In: Sander Uwe., von Gross Frederike., Hugger Kai-Uwe (Hrsg.), *Handbuch Medienpädagogik.* VS Verlag für Sozialwissenschaften, S. 87–92. https://doi.org/10.1007/978-3-531-91158-8_9

Borhrmann, T, Reichelt, M. & Veith, W. (2018). (Hrsg.), *Angewandte Ethik und Film.* VS für Sozialwissenschaften.

Bohrmann, T. (2018). Einführung in die ethische Filmanalyse. In Borhrmann, Thomas; Reichelt, Matthias & Veith, Werner (Hrsg.), *Angewandte Ethik und Film.* Wiesbaden: VS Verlag für Sozialwissenschaften, S. IX-XVIII.

Borhrmann, T. Reichelt, M. & Veith, W. (2018). Einleitung: Warum der Film auch ein Gegenstand der Ethik ist. In Borhrmann, Thomas; Reichelt, Matthias & Veith, Werner (Hrsg.), *Angewandte Ethik und Film.* VS Verlag für Sozialwissenschaften, S. 37–57.

Bosshart, L. (2007). Information und/oder Unterhaltung? In: Scholl, Armin, Renger, Rudi & Blöbaum, Bernd (Hrsg.), *Journalismus und* Unterhaltung. Theoretische Ansätze und empirische Befunde. VS Verlag für Sozialwissenschaften, S. 17–29.

Buxmann P., Schmidt H. (2019). Ökonomische Effekte der Künstlichen Intelligenz. In P. Buxmann & H. Schmidt (Hrsg.), *Künstliche Intelligenz.* Springer Gabler. https://doi.org/10.1007/978-3-662-57568-0_2

Cave, S., & Dihal, K. (2019). Hopes and fears for intelligent machines in fiction and reality. *Nature Machine Intelligence*, 1(2), 74–78. https://doi.org/10.1038/s42256-019-0020-9

Chun, W. H. K. (2005). On software, or the persistence of visual knowledge. *Grey Room*, *18,* 26–51.

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence.* Yale University Press.

Damberger, T. & Hebert, E. (2017). Is Pedagogy Transhuman? Reflections on the Relationship Between Pedagogy and Transhumanism. *Studia Paedagogica Ignatiana*, 20(1), 41–56.

De Witt, C., & Leineweber, C. (2020). Zur Bedeutung des Nichtwissens und die Suche nach Problemlösungen: Bildungstheoretische Überlegungen zur Künstlichen Intelligenz. *MedienPädagogik: Zeitschrift Für Theorie Und Praxis Der Medienbildung*, 39, 32–47. https://doi.org/10.21240/mpaed/39/2020.12.03.X

Dotzler, B. (2018). „Down-to-earth resolutions“. Erinnerungen an die KI als eine „häretische Theorie“. In C. Engemann & A. Sudmann (Hrsg.), *Machine Learning Medien, Infrastrukturen und Technologien der Künstlichen Intelligenz*, S. 39–54. transcript.

Ess, C. (2013). *Digital Media Ethics.* Polity.

Floridi, L., & Cowls, J. (2021). A unified framework of five principles for AI in society. In L. Floridi (Hrsg), *Ethics, Governance, and Policies in Artificial Intelligence*, S. 5–17. Springer VS. https://link.springer.com/chapter/10.1007/978-3-030-81907-1_2

Fry, H. (2019). *Hello world: Was Algorithmen können und wie sie unser Leben verändern* (S. Schmid, Übers.). C.H. Beck.

Gerrig, R. J., & Zimbardo, P. G. (2008). *Psychologie* (18., aktualisierte Auflage). Pearson Studium.

Giesinger, J. (2020). Autonomie. In G. Weiß & J. Zirfas (Hrsg.), *Handbuch Bildungs- und Erziehungsphilosophie*, S. 235–244. Springer Fachmedien. https://doi.org/10.1007/978-3-658-19004-0_21

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). XAI—Explainable artificial intelligence. *Science Robotics*, *4*(37), eaay7120. https://doi.org/10.1126/scirobotics.aay7120

Hartong, S. (2019). Learning Analytics und Big Data in der Bildung. Zur notwendigen Entwicklung eines datenpolitischen Alternativprogramms. GEW. Online: https://www.gew.de/index.php%3FeID%3DdumpFile%26t%3Df%26f%3D91791%26token%3D702ec8d5f9770206a4aa8a1079750ec9021b90bf%26sdownload%3D%26n%3DLearning-analytics-2019-web-IVZ.pdf [Zugriff: 6. 6. 2021]

Hausmanninger, T. (1993). *Kritik der medienethischen Vernunft. Die ethische Diskussion über den Film in Deutschland im 20. Jahrhundert.* Fink.

Klein, B. (2020). *Hilfsmittel, Assistive Technologien und Robotik: Selbstständigkeit und Lebensqualität im Alter erhalten.* Kohlhammer.

Krotkov, E., Hackett, D., Jackel, L., Perschbacher, M., Pippine, J., Strauss, J., Pratt, G., & Orlowski, C. (2017). The DARPA Robotics Challenge Finals: Results and Perspectives: The DARPA Robotics Challenge Finals. *Journal of Field Robotics*, 34(2), 229–240. https://doi.org/10.1002/rob.21683

Lafrance, A. (2014). Why people name their machines. *The Atlantic.* URL: http://www.theatlantic.com/technology/archive/2014/06/why-people-give-human-names-to-machines/373219/ [Zugriff: 14. 1. 2022]

Larson, E. J. (2021). *The myth of artificial intelligence: Why computers can't think the way we do.* The Belknap Press of Harvard University Press.

Lee, B. (2015). Ex Machina stunt at SXSW has users falling for a robot on Tinder. The Guardian. URL: https://www.theguardian.com/film/2015/mar/16/ex-machina-stunt-sxsw-users-falling-for-robot-tinder. [Zugriff: 2. 7. 2021]

Lingel, J, & Crawford, K. (2020) „Alexa, tell me about your mother“: The history of the secretary and the end of secrecy. *Catalyst: Feminism, Theory, Technoscience*, 6(1), 1–25.

Mainzer, K. (2019). Künstliche Intelligenz – Wann übernehmen die Maschinen? (2., erweiterte Auflage). Springer VS.

Marotzki, W. & Jörissen, B. (2009) Medienbildung – eine Einführung. Klinkhardt UTB.

Plato (2007). *The Republic.* (2nd Edition; trans. by Lee, D.) Penguin Classics.

Russell, S. J., & Norvig, P. (2021). Artificial intelligence: A modern approach (Fourth edition). Pearson.

Sesink, W. (1993). Menschliche und künstliche Intelligenz: Der kleine Unterschied. Klett-Cotta.
Silver, D., Singh, S., Precup, D., & Sutton, R. S. (2021). Reward is enough. *Artificial Intelligence*, *299*, 103535. https://doi.org/10.1016/j.artint.2021.103535
Simanowski, R. (2020). *Todesalgorithmus: Das Dilemma der künstlichen Intelligenz.* Passagen Verlag.
Stammberger, B. (2014). Haare als Symptom. Diskurse über Weiblichkeit, Schönheit und Identität. *Body Politics* 2 (2), 431–461.
Steiner, P. (1993). On the internet, nobody knows you're a dog. The New Yorker. 5. 7. 1993.
Thimm, C., & Bächle, T. C. (2019). Autonomie der Technologie und autonome Systeme als ethische Herausforderung. In M. Rath, F. Krotz, & M. Karmasin (Hrsg.), *Maschinenethik*, S. 73–87. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-21083-0_5
Turing, A. M. (1950). I.—COMPUTING MACHINERY AND INTELLIGENCE. Mind, LIX(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433
Wendt, D. H. & Jung, C. (2020). Europäischer Rechtsrahmen für Künstliche Intelligenz (Teil I), LR 2020, 34–41. URL: https://legal-revolution.com/images/pdf/Europischer_Rechtsrahmen_fr_Knstliche_Intelligenz_Teil_I.pdf [Zugriff: 15. 1. 2022]
Williamson, B. (2017). *Big Data in Education.* Sage Publishing.
Wolf, F. (2021). „Das Private ist digital und öffentlich": Betrachtungen über das Verhältnis zwischen Privatheit und Öffentlichkeit am Beispiel der Smartphonenutzung im Alter. *Psychotherapie im Alter*, *18*(1), 71–82. https://doi.org/10.30820/1613-2637-2021-1-71
Zweig, K. A. (2019). *Ein Algorithmus hat kein Taktgefühl: Wo künstliche Intelligenz sich irrt, warum uns das betrifft und was wir dagegen tun können* (Originalausgabe). Heyne.

Künstliche Intelligenz und Fairness im Bildungskontext

Katharina Simbeck

1 Maschinelles Lernen und Künstliche Intelligenz

Anwendungen der Künstlichen Intelligenz (KI) versuchen komplexe Aufgaben zu lösen, die bisher nur von Menschen gelöst werden konnten. Die heute übliche Form der künstlichen Intelligenz ist das maschinelle Lernen (ML), mit welchem Zusammenhänge in Trainingsdaten erkannt und auf neue Daten übertragen werden. Maschinelles Lernen setzt deshalb die Verfügbarkeit von großen Datenmengen voraus. Für die Verfahren des überwachten Lernens muss die zu erkennende Eigenschaft schon in den Trainingsdaten markiert sein. Sollen beispielsweise Aufgabenstellungen im Hinblick auf ihre Schwierigkeit klassifiziert werden, müssen die Trainingsdaten Variablen (im maschinellen Lernen als Features bezeichnet) enthalten, die Aussagen über die Schwierigkeit zulassen, bspw. die Anzahl richtig/falsch gelöster Bearbeitungen. Beim unüberwachten Lernen dagegen werden Ähnlichkeiten in Datensätzen gesucht. So können bspw. Lernende an Hand von Logdaten aus Lernmanagementsystemen in Gruppen eingeteilt werden (Clustering).

Interessanterweise lassen sich die Aufgaben des maschinellen Lernens (z. B. Klassifikation, Clustering) mit ganz unterschiedlichen Methoden umsetzen. Häufig probiert man verschiedene Methoden aus, um die passende für den Anwendungsfall zu finden. Bekannte Methoden sind beispielsweise Entscheidungsbäume, Support Vector Machines (SVMs), Bayes'sche Netze und Neuronale Netze. Entscheidungsbäume stellen eine geordnete Reihe von Bedingungen dar, welche eine Eingabe der passenden Ausgabe zuordnen. Neuronale Netze ahmen die Funktionsweise von biologischen neuronalen Netzen durch gewichtete Matrizen-Multiplikationen nach, sie sind in der Lage, erstaunlich komplexe Aufgaben zu lösen.

Vor der Anwendung der ML-Verfahren werden die zur Verfügung stehenden Daten in Trainings- und Testdaten unterteilt. Die Trainingsdaten werden genutzt, um das ML-Modell zu trainieren. Mithilfe der Testdaten wird geprüft, ob das Modell auch auf anderen Daten funktioniert. So soll ein Bilderkennungssystem ja z. B. Äpfel im Allgemeinen erkennen und nicht nur die Apfelbilder wiedererkennen, die es schon „gesehen“ hat. Letzteren Fall bezeichnet man als Überanpassung (overfitting).

2 Künstliche Intelligenz im Bildungskontext

Durch die Verwendung von digitalen Lernmaterialien und Lernmanagementsystemen entstehen immer mehr Daten über Lernende, Lehrende und Lernprozesse, welche ausgewertet und genutzt werden können. Learning Analytics beinhaltet die „Messung, Sammlung, Analyse und das Reporting von Daten über Lernende und ihre Kontexte, für die Zwecke des Verstehens und Optimierens des Lernens und der Umgebung, in der es auftritt“ (Siemens 2013). Nach Grandl et al. (2018) verfolgt Learning Analytics fünf Hauptziele:

– Vorhersage des Lernerfolgs,
– Empfehlungen für angemessene Interventionen,
– Anpassung der Lernumgebung basierend auf der Lernaktivität,
– Die Lernenden sollen in die Lage versetzt werden, über ihren Lernprozess zu reflektieren,
– Evaluation von Methoden oder Lerninhalte.

Für Vorhersagen, Empfehlungen oder Anpassungen der Lernumgebung ist es üblich, Verfahren des maschinellen Lernens auf Lerndaten anzuwenden (Ochoa & Merceron 2018).

Für KI-Anwendungen im Bildungswesen können verschiedene Arten von Daten verwendet werden, bspw. Textdaten, soziodemografische Daten oder Lernprozessdaten, die durch die Anwendung digitaler Lernplattformen entstehen. Wenn digitale Texte vorliegen, z. B. Aufsätze oder Aufgabenstellungen, können KI-Verfahren (Natural Language Processing) angewandt werden, um die syntaktische Komplexität oder die lexikalische Vielfalt zu beurteilen und Rückschlüsse auf die Schwierigkeit der Aufgabenstellung oder die Qualität des Lernergebnisses zu ziehen (McNamara et al. 2017).

Auch die Vorhersage des Lernerfolgs ist häufig Gegenstand von Learning Analytics (Yadav et al. 2011, Shahiri et al. 2015, Chen et al. 2020), insbesondere um individuelle Unterstützungsangebote zu unterbreiten. Der Einsatz von KI im Bildungsbereich öffnet darüber hinaus Chancen zur Entlastung der Lehrenden und zur Verbesserung der Lehrqualität. Die Korrektur von Schül-

erarbeiten bspw. gehört zu den belastendsten Aufgaben der Lehrkräfte (forsa 2016). Für viele schulische Lerninhalte ist aber ein sofortiges und häufiges Feedback entscheidend für den Lernerfolg (Anderson & Funke 2013, Fengler 2017). Die Häufigkeit und Aktualität des Feedbacks an die Lernenden wird wiederum durch den manuellen Korrekturaufwand der Lehrkräfte eingeschränkt. Digital und KI-gestütztes Lernen dagegen ermöglicht eine automatisierte Differenzierung der Aufgabenschwierigkeit, schnelles und passendes Feedback und damit die Möglichkeit, auch bei Lernlücken die Motivation länger aufrecht zu halten (Stöcklin 2018).

In heterogenen Klassen ist es nur selten möglich, individuell angemessene und geeignete Aufgaben zu stellen, sowohl in Bezug auf das Niveau als auch auf die Art der Aufgabe. Dies kann dazu führen, dass die Erfolgserlebnisse der Schülerinnen und Schüler geringer werden und es zu einem Verlust an Selbstvertrauen und Motivation kommt (Köller et al. 2006, Marsh et al. 2005). Ein KI-gestütztes Lernsystem könnte eine Assistenzfunktion für die Lehrkraft übernehmen und die Lernenden individuell in ihrer Entwicklung begleiten. Ein besseres Verständnis von Lernprozessen kann auch Hinweise zur Weiterentwicklung von Curricula geben (Rzepka et al. 2021).

3 Diskriminierung durch KI

Die öffentliche Debatte über algorithmische Fairness, Verantwortung, Transparenz und Ethik der Künstlichen Intelligenz begann mit einem Artikel, der rassistische Vorhersagen zur Rückfälligkeit von Straftätern kritisierte (Angwin et al. 2016).

Da beim maschinellen Lernen Zusammenhänge aus großen Datensätzen erkannt und abstrahiert werden, ist es nicht verwunderlich, dass die vorherrschenden gesellschaftlichen Stereotypen in diesen Datensätzen enthalten sind und in die Modelle übernommen werden. Dies trifft auch auf KI-Anwendungen im Bildungsbereich zu (Riazy & Simbeck 2019, Riazy et al. 2020). KI-Systeme können systematisch unterschiedliche Ergebnisse z. B. für Frauen, Ältere oder Personen mit einer bestimmten Postleitzahl liefern und dadurch diskriminierend wirken (Feldman et al. 2015). Dieser Effekt wird als disparate impact bezeichnet (ebd.). Dies ist selbst dann möglich, wenn diese Merkmale (z. B. Geschlecht) nicht im Datensatz enthalten sind. Diese indirekte Diskriminierung entsteht dann, wenn die Merkmale mit anderen stark korrelieren – z. B. werden Frauencolleges nur von Frauen besucht.

Die Messung der Fairness oder Diskriminierungs-Freiheit von KI-Systemen ist komplex. Es wurden eine ganze Reihe von Fairness-Kennzahlen entwickelt (Dwork et al. 2012, Hardt et al. 2016, Verma & Rubin 2018), die unter-

schiedliche Aspekte von Fairness betrachten. Gebräuchliche Fairness-Maße beschreiben zum einen die Ausgewogenheit der Trainingsdaten (Kamishima et al. 2012, Williamson & Menon 2019) oder messen inwiefern falsch positive und falsch negative Ergebnisse in einzelnen Gruppen häufiger auftreten (Kusner et al. 2017, Hardt et al. 2016).

Um die Fairness von KI-Anwendungen zu verbessern, können die verwendeten Trainingsdaten optimiert (Calmon et al. 2017, Adler et al. 2016) angepasste algorithmische Verfahren verwendet werden (Zafar et al. 2015, Kamishima et al. 2012, Williamson & Menon 2019, Calders & Verwer 2010) oder die Diskriminierungsfreiheit kann nach der Anwendung des ML-Verfahrens umgesetzt werden (Hardt et al. 2016, Friedler et al. 2019, Feldman et al. 2015, Calders & Verwer 2010). Viele Autor*innen suchen einen Kompromiss zwischen Fairness und Genauigkeit eines Algorithmus (Friedler et al. 2019, Williamson & Menon 2019).

Die Fairness des maschinellen Lernens hängt aber wesentlich vom Kontext der zu treffenden Entscheidung ab (Chouldechova & Roth 2018, O'Neil 2017), denn die Wirkung eines KI-Systems entsteht im Zusammenspiel von verwendeten Daten, Algorithmen und dem Verhalten der Nutzer*innen. Dieses Verhalten wiederum wird von den Erwartungen und Annahmen der Nutzer*innen über die Funktionsweise des Systems sowie ihren persönlichen Präferenzen bestimmt. Deshalb reicht es nicht aus, Kennzahlen zur Messung der Fairness des KI-Modells zu verwenden, sondern die komplexere Interaktionen zwischen Nutzer*innen und System muss auf Systemebene betrachtet werden (Holstein et al. 2019). Machine-Learning-Systeme sollten also nicht nur „in vitro", sondern auch in den sie einbettenden soziotechnischen Kontexten „in vivo" betrachtet werden (Veale et al. 2018).

4 KI im Bildungsbereich, Ethische Aspekte

Der subjektiven Wahrnehmung von Fairness seitens der Lehrenden und Lernenden kommt damit eine entscheidende Rolle bei der Bewertung von KI-Anwendungen im Bildungsbereich zu, die nicht allein rein formal-analytisch oder auf Basis von Kennzahlen erfolgen sollte.

Die Fairness und Validität von KI-Systemen im Lernkontext sind von entscheidender Bedeutung, da sich die Vorhersagen der Schülerleistungen auf den Erfolg der Lernenden auswirken (Rosenthal & Rubin 1978) und somit Ungleichheiten aufrechterhalten oder sogar verstärken können (O'Neil 2017). Lernende in adaptiven Lernumgeben, denen fälschlich eine niedrige Leistung vorhergesagt wird und die deshalb einfachere Aufgaben erhalten, verlieren die Chance höherwertige Kompetenzstufen zu erreichen. Es gibt erst wenige An-

sätze, die beschreiben, wie Learning Analytics Systeme systematisch auf Fairness geprüft werden können (Gardner et al. 2019, Mayfield et al. 2019, Porayska-Pomsta & Rajendran 2019).

In den letzten Jahren wurden dafür mehrere detaillierte Ethik-Richtlinien für Learning Analytics veröffentlicht (Yun et al. 2019, Prinsloo & Slade 2017, Drachsler & Greller 2016, Welsh & McKinney 2015, Sclater & Bailey 2018). Die Richtlinieren betonen in der Regel die Notwendigkeit von Zustimmung zur Datennutzung, Transparenz und Datenschutz (Prinsloo & Slade 2017, Drachsler & Greller 2016, Ferguson et al. 2016, Sclater & Bailey 2018, Sclater 2016). In ihrem wegweisenden Beitrag fordern Prinsloo und Slade (2017) darüber hinaus, dass Lernende als Agenten ihres Lernens und nicht nur als Empfangende von Lerninterventionen betrachtet werden. Daraus folgt, dass Lernenden kein „magisches“ KI-System vorgesetzt werden sollte, sondern sie müssen die Möglichkeit haben, den Zusammenhang zwischen Input- und Outputvariablen zumindest in groben Zügen zu verstehen. Darüber hinaus weisen Prinsloo und Slade (2017) auf die Dynamik des Lernprozesses hin, Lernende dürfen nicht mit veralteten Daten des eigenen Ichs oder daraus abgeleiteten Interventionen konfrontiert werden. Übertragen auf KI-gestützt individualisierte Lernpfade bedeutet dies, dass Lernende die Möglichkeit haben müssen, von dem vorgeschlagenen Lernpfad abzuweichen. Auch umfassende digitale Lernplattformen können Lernprozesse nur ausschnittsweise digital erfassen – die Lernenden könnten zusätzlich Bücher lesen oder den Lerninhalt durch analoge Konversationen rekapitulieren. Wie auch von Prinsloo und Slade (2017) gefordert, darf die Komplexität des Lernens nicht auf die digital verfügbaren Daten reduziert werden.

5 Konsequenzen für den Einsatz von KI im Bildungsbereich

Es ist nicht nötig und auch nicht möglich, dass Lehrende und Lernende im Bildungsbereich eingesetzte KI-Systeme im Detail verstehen, nachvollziehen oder überprüfen können: Die Systeme verwenden große Datenmengen, die sich kontinuierlich ändern. Die dahinterliegenden Modelle und Algorithmen sind auch für Expert*innen nur mit großem Zeitaufwand nachvollziehbar.

Für viele Anwendungsgebiete ist es jedoch notwendig, die Ausgaben des Systems, d. h. die Einflussfaktoren auf die Systemergebnisse verstehen zu können, um diese zu akzeptieren oder ggf. auch korrigieren zu können. Dieses Ziel kann auf der Ebene der Benutzerschnittstelle teilweise erreicht werden (vgl. bspw. Scheers & Laet 2021):

- durch die Darstellung von Variablen, die das Ergebnis beeinflussen,
- durch die graphische oder textuelle Darstellung der Unschärfe von Ergebnissen
- durch die Möglichkeit, Ergebnisse durch Änderung von Variablen zu beeinflussen,
- durch die Darstellung von Kennzahlen zur Fairness,
- durch kontrafaktische Erklärungen („ohne diesen Fehler hättest Du eine schwerere Aufgabe bekommen").

Die Digitalisierung der Lernprozesse, weit über die Bereitstellung von digitalem Lernmaterial hinaus, wurde durch die Covid-19-Pandemie unumkehrbar beschleunigt. Digitale Lernprozesse generieren Daten über Lernende, Lehrende und den Lernprozess selbst. Darüber hinaus bieten digitale Lernplattformen die Möglichkeit durch formative und summative Lernstandsmessungen individuelle Kompetenzniveaus kontinuierlich zu bestimmen. Diese Daten können wiederum genutzt werden, um Lernprozesse besser zu verstehen und individualisiert zu begleiten. Individuell angepasste Schwierigkeiten und schnelles, passendes Feedback in digitalen Lernumgebungen können die Motivation der Lernenden aufrechterhalten und Lehrende entlasten.

Lehrende, Lernende und Entscheidungsträger*innen im Bildungssystem müssen sich jedoch bewusst sein, dass den Chancen die sich durch datengetriebene Verfahren ergeben eine Reihe von Herausforderungen gegenüberstehen. Diese Herausforderungen betreffen zum einen Datenschutz, Transparenz und die informierte Möglichkeit, auf die Datenverarbeitung zu verzichten. Zum anderen ergeben sich aber auch Herausforderungen in Bezug auf die Transparenz und Diskriminierungsfreiheit der eingesetzten Systeme:

- Sind die Prognosen systematisch weniger zuverlässig für Lernende mit Migrationshintergrund?
- Bekommen Jungen systematisch herausfordernde Matheaufgaben angeboten als Mädchen?
- Werden die Anforderungen von unterproportional vertretenen Gruppen (z. B. Lernende mit Behinderung, mit anderer Muttersprache) ausreichend berücksichtigt?

Da sich die Wirkung des KI-Systems durch das Zusammenwirken von Daten und Modell im Lernprozess ergibt, ist eine Beurteilung der Fairness allein durch Analyse von Daten oder Algorithmus nicht möglich. Lernenden und Lehrenden muss deshalb über die Benutzerschnittstelle des Systems die Möglichkeit gegeben werden, Annahmen zur Funktionsweise kontinuierlich zu verifizieren.

Literaturverzeichnis

Adler, P., Falk, C., Friedler, S. A., Rybeck, G., Scheidegger, C., Smith, B., & Venkatasubramanian, S. (2016). Auditing Black-box Models for Indirect Influence. https://doi.org/10.48550/ARXIV.1602.07043

Anderson, J. R. & Funke, J. (Hrsg.) (2013). *Kognitive Psychologie. 7., erw. und überarb., neu gestaltete Aufl. Lehrbuch.* Springer VS.

Angwin, J., Larson, J., Mattu, S. (2016). Machine bias: There's software used across the country to predict future criminals and it's biased against blacks. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [Zugriff: 3. 9. 2021]

Calders, T., Verwer, S. (2010). Three naive Bayes approaches for discrimination-free classification. *Data Mining and Knowledge Discovery* 21(2), 277–292.

Calmon, F. P., Wei, D., Ramamurthy, K. N., & Varshney, K. R. (2017). *Optimized Data Pre-Processing for Discrimination Prevention.* https://doi.org/10.48550/ARXIV.1704.03354

Chen, Y., Zheng, Q., Ji, S., Tian, F., Zhu, H., & Liu, M. (2020). Identifying at-risk students based on the phased prediction model. *Knowledge and Information Systems*, *62*(3), 987–1003. https://doi.org/10.1007/s10115-019-01374-x

Chouldechova, A. & Roth, A. (2018). The Frontiers of Fairness in Machine Learning.

Drachsler, H., & Greller, W. (2016). Privacy and analytics: It's a DELICATE issue a checklist for trusted learning analytics. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge – LAK '16*, 89–98. https://doi.org/10.1145/2883851.2883893

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference on – ITCS '12*, 214–226. https://doi.org/10.1145/2090236.2090255

Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., & Venkatasubramanian, S. (2015). Certifying and Removing Disparate Impact. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 259–268. https://doi.org/10.1145/2783258.2783311

Fengler, J. (2017). *Feedback geben. Strategien und Übungen.* 5., überarbeitete und erweiterte Auflage 2017. Beltz.

Ferguson, R., Hoel, T., Scheffel, M., & Drachsler, H. (2016). Guest Editorial: Ethics and Privacy in Learning Analytics. *Journal of Learning Analytics*, *3*(1). https://doi.org/10.18608/jla.2016.31.2

forsa (2016). Zufriedenheit im Lehrberuf. Ergebnisse einer repräsentativen Lehrerbefragung.

Friedler, S. A., Scheidegger, C., Venkatasubramanian, S., Choudhary, S., Hamilton, E. P., & Roth, D. (2019). A comparative study of fairness-enhancing interventions in machine learning. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 329–338. https://doi.org/10.1145/3287560.3287589

Gardner, J., Brooks, C., & Baker, R. (2019). Evaluating the Fairness of Predictive Student Models Through Slicing Analysis. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 225–234. https://doi.org/10.1145/3303772.3303791

Grandl, M., Taraghi, B., Ebner, M., Leitner, P., & Ebner, M. (2018). Learning Analytics. In P. Arnold, L. Kilian, A. M. Thillosen & G. M. Zimmer (Hrsg.), *Handbuch E-*

Learning. Lehren und Lernen mit digitalen Medien. 5. Auflage. utb Pädagogik, Band 4965. W. Bertelsmann Verlag, S. 1–16.

Hardt, M., Price, E., & Srebro, N. (2016). *Equality of Opportunity in Supervised Learning.* https://doi.org/10.48550/ARXIV.1610.02413

Holstein, K., Wortman Vaughan, J., Daumé, H., Dudik, M., & Wallach, H. (2019). Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need? *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–16. https://doi.org/10.1145/3290605.3300830

Kamishima, T., Akaho, S., Asoh, H., & Sakuma, J. (2012). Fairness-Aware Classifier with Prejudice Remover Regularizer. In P. A. Flach, T. De Bie, & N. Cristianini (Hrsg.), *Machine Learning and Knowledge Discovery in Databases* (Bd. 7524). Springer, S. 35–50. https://doi.org/10.1007/978-3-642-33486-3_3

Köller, O., Trautwein, U., Lüdtke, O., & Baumert, J. (2006). Zum Zusammenspiel von schulischer Leistung, Selbstkonzept und Interesse in der gymnasialen Oberstufe. *Zeitschrift für Pädagogische Psychologie*, *20*(1/2), 27–39. https://doi.org/10.1024/1010-0652.20.12.27

Kusner, M. J., Loftus, J. R., Russell, C., & Silva, R. (2017). *Counterfactual Fairness.* https://doi.org/10.48550/ARXIV.1703.06856

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic Self-Concept, Interest, Grades, and Standardized Test Scores: Reciprocal Effects Models of Causal Ordering. *Child Development*, *76*(2), 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x

Mayfield, E., Madaio, M., Prabhumoye, S., Gerritsen, D., McLaughlin, B., Dixon-Román, E., & Black, A. W. (2019). Equity Beyond Bias in Language Technologies for Education. *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, 444–460. https://doi.org/10.18653/v1/W19-4446

McNamara, D. S., Allen, L. K., Crossley, S. A., Dascalu, M., & Perret, C. A. (2017). Natural Language Processing and Learning Analytics. In C. Lang, G. Siemens, A. Wise, & D. Gasevic (Hrsg.), *Handbook of Learning Analytics.* Society for Learning Analytics Research (SoLAR). S. 93–104. https://doi.org/10.18608/hla17.008

Ochoa, X. & Merceron, A. (2018). Quantitative and Qualitative Analysis of the Learning Analytics and Knowledge Conference 2018. *Journal of Learning Analytics* 5(3), 154–166.

O'Neil, C. (2017). *Weapons of math destruction. How big data increases inequality and threatens democracy. First paperback edition.* Broadway Books.

Porayska-Pomsta, K. & Rajendran, G. (2019). Accountability in Human and Artificial Intelligence Decision-Making as the Basis for Diversity and Educational Inclusion. In J. Knox, Y. Wang & M. Gallagher (Hrsg.), *Artificial Intelligence and Inclusive Education: Speculative Futures and Emerging Practices.* Springer Singapore, S. 39–59.

Prinsloo, P. & Slade, S. (2017). Ethics and learning analytics: Charting the (un)charted. C. Lang, C. G. Siemens, A. Wise & D. Gašević (Hrsg.), *Handbook of learning analytics. First edition.* Solar Society for Learning Analytics Research, S. 49–57.

Riazy, S. & Simbeck, K. (2019). Predictive Algorithms in Learning Analytics and their Fairness. In N. Pinkwart & J. Konert (Hrsg.), *DeLFI 2019 – Die 17. Fachtagung Bildungstechnologien.* Gesellschaft für Informatik e.V, S. 223–228.

Riazy, S., Simbeck, K. & Schreck, V. (2020). Fairness in Learning Analytics: Student At-risk Prediction in Virtual Learning Environments. In *Proceedings of the 12th*

International Conference on Computer Supported Education. SCITEPRESS – Science and Technology Publications, S. 15–25.

Rosenthal, R. & Rubin, D. B. (1978). Interpersonal expectancy effects: the first 345 studies. *Behavioral and Brain Sciences* 1(3), 377–386.

Rzepka, N., Müller, H.-G. & Simbeck, K. (2021). What you apply is not what you learn! Examining students' strategies in German capitalization tasks. In *Educational Data Mining.* URL: https://educationaldatamining.org/EDM2021/virtual/static/pdf/EDM21_paper_68.pdf [Zugriff: 23. 8. 2021]

Scheers, H. & Laet, T. de (2021). Interactive and Explainable Advising Dashboard Opens the Black Box of Student Success Prediction. In T. de Laet, R. Klemke, C. Alario-Hoyos, I. Hilliger, A. Ortega-Arranz (Hrsg.), *Technology-Enhanced Learning for a Free, Safe, and Sustainable World. Lecture Notes in Computer Science*, Band 12884. Springer International Publishing, S. 52–66.

Sclater, N. (2016). Developing a Code of Practice for Learning Analytics. *Journal of Learning Analytics* 3(1), 16–42.

Sclater, N. & Bailey, P. (2018). Code of practice for learning analytics. https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics [Zugriff: 14. 9. 2021].

Shahiri, A. M., Husain, W. & Rashid, N. A. (2015). A Review on Predicting Student's Performance Using Data Mining Techniques. *Procedia Computer Science* 72, 414–422.

Siemens, G. (2013). Learning Analytics: The ergemence of a discipline. *American Behavioral Scientist* 57(10), 1380–1400.

Stöcklin, N. (2018). *Computerunterstützte Gamifizierung in der Sekundarstufe I. Konzeption und Erforschung von Maßnahmen zur Einbindung spielerischer Elemente in Lernsettings.* Dissertation. Pädagogische Hochschule Heidelberg/Fakultät für Natur- und Gesellschaftswissenschaften (Fak. III).

Veale, M., van Kleek, M. & Binns, R. (2018). Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making. In: Association for Computing Machinery (Hrsg.), *CHI'18 Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* The Association for Computing Machinery, S. 1–14.

Verma, S. & Rubin, J. (2018). Fairness definitions explained. In Y. Brun, Y., B. Johnson & A. Meliou (Hrsg.), *Proceedings of the International Workshop on Software Fairness – FairWare '18.* New York, New York, USA: ACM Press, S. 1–7.

Welsh, S. & McKinney, S. (2015). Clearing the Fog: A Learning Analytics Code of Practice. In Australasian Society for Computers in Learning in Tertiary Education (Hrsg.), *Globally connected, digitally enabled*, S. 588–592.

Williamson, R. & Menon, A. (2019). Fairness risk measures. In: JMLR Workshop and Conference Proceedings (Hrsg.), *Proceedings of the 36th International Conference on Machine Learning*, ICML 2019, 9–15 June 2019, Long Beach, California, USA. Proceedings of Machine Learning Research, Band 97, S. 6786–6797.

Yadav, S. K., Bharadwaj, B., & Pal, S. (2012). Data Mining Applications: A comparative Study for Predicting Student's performance. *International Journal of Innovative Technology and Creative Engineering* 1(12), https://doi.org/10.48550/ARXIV.1202.4815

Yun, H., Riazy, S., Fortenbacher, A. & Simbeck, K. (2019). Code of practice for sensor-based learning. In N. Pinkwart & J. Konert (Hrsg.), *DeLFI 2019 – Die 17. Fachtagung Bildungstechnologien.* Gesellschaft für Informatik e.V.

Zafar, M. B., Valera, I., Rodriguez, M. G., & Gummadi, K. P. (2015). *Fairness Constraints: Mechanisms for Fair Classification.* https://doi.org/10.48550/ARXIV.1507.05259

Autonomie in digitalen Kontexten geschlechterkritisch vermessen

Andrea Wolffram

Einleitung

Vor einigen Jahren wurden in den Debatten über die Digitalisierung der Lebenswelt, und hier insbesondere der Arbeitswelt, Geschlechterperspektiven noch weitgehend ausgeblendet. Eine solche Perspektive macht jedoch insbesondere unter dem Gesichtspunkt einer gleichberechtigen Autonomie auf notwendig zu differenzierende Aspekte in den jeweiligen gesellschaftlichen Feldern aufmerksam. Dies hat sich in jüngster Zeit verändert, und es liegen eine Reihe von Analysen zu unterschiedlichen Digitalisierungsbereichen vor (u. a. Carstensen 2015, 2019; Kutzner &Schnier 2017, Prietl 2019). Nicht zuletzt das Anfang 2021 veröffentlichte Gutachten für den 3. Gleichstellungsbericht der Bundesregierung (Sachverständigenkommission für den Dritten Gleichstellungsbericht der Bundesregierung 2021, nachfolgend 3. GSB 2021) setzte hier seinen Schwerpunkt. Das Anliegen meines Beitrages ist es, vor diesem Hintergrund nicht nur verschiedene Facetten der Digitalisierung aus Geschlechterperspektive zu beleuchten, was vermutlich nur wenig Neues den bestehenden Analysen hinzufügen würde. Vielmehr ist mein Ansinnen, dies unter einer Autonomie-Perspektive vorzunehmen, insofern diese aus geschlechterkritischer bzw. theoretischer Sicht ein bedeutsames Potenzial aufweist, ungleiche Geschlechterverhältnisse mit einhergehender Diskriminierung auf den sozialen Gegenstand „Digitalisierung“ zu beziehen.

Autonomie im digitalen Zeitalter geschlechtertheoretisch reflektiert

Aus Geschlechterperspektive besteht ein Zugang zur Reflexion des Begriffes Autonomie in seiner verkürzten Geltung auf das Subjekt in seinem Zuschnitt einer „hegemoniale[n] Männlichkeit“ (Connell 1999). Die Frage, wie Autonomie aus geschlechtertheoretischer Perspektive gedacht werden kann, setzt so zunächst an dieser damit einhergehenden grundlegenden Kritik an. Aus diesem eingegrenzten Geltungsbereich leiten Baer und Saksofsky (2018, S. 13) vier Stränge der Kritik ab, die jeweils auf Fragen der Benachteiligung im Geschlechterverhältnis rekurrieren. Es geht „um die Ausgrenzung von Frauen und die – insofern nicht zufällige – Ausblendung von Frauen“ aus Sphären der Öffentlichkeit (1) und damit einhergehend ihre Zuweisung zum Privaten über Schwangerschaft und Mutterschaft bzw. care work (2). Daran knüpft sich die Benachteiligung von Frauen in Familie, Erwerbsarbeit und sozialer Sicherung an. Es geht ferner darum, wie die „Erfahrung der Opfer von Gewalt berücksichtigt werden [… und] Gewalt als eine Form geschlechtsbezogener Diskriminierung und damit im Kontext gesellschaftlicher Strukturen zu verstehen“ (ebd., S. 14) ist (3). Damit ist nicht nur Freiheit und Selbstbestimmung im Sinne von Autonomie adressiert, sondern auch Gleichheit. Hier gilt es, (4) ein „formelles Gleichheitsverständnis als de facto diskriminierend zu entlarven und Gleichheit als Recht auf materielle Gleichberechtigung und den daraus abzuleitenden Diskriminierungsschutz zu konzeptionieren“ (ebd., S. 14).

Juristinnen modifizierten Gleichheit somit „als Gleichberechtigung, Gleichstellung, Chancen- oder Ergebnisgleichheit, als Dominierungs- oder Hierarchisierungsverbot“ (ebd. 14). Dies ging einher mit der Suche nach „Ansätzen, die jede Bevormundung durch Angleichung ebenso verhindern wie sie einem Leerlaufen nur formaler oder symmetrischer Gleichheitsversprechen entgegentreten“ (ebd., S. 14). Entscheidend ist hierbei, dass der Begriff der Autonomie eine Schärfung erfährt, insofern Autonomie nicht nur Freiheit von Einzelnen umfasst, sondern Freiheit unter Gleichen und sich damit Gleichheit vor allem auf gleichberechtigte Autonomie bezieht.[1]

Die vier Stränge der Kritik am Begriff der Autonomie und die Einforderung von Autonomie im Zuschnitt von Freiheit, Selbstbestimmung und Gleichheit behalten ihre Gültigkeit auch mit Blick auf die „Digitalisierung“ und können wie folgt übersetzt werden:

– Ausgrenzung/Ausblendung aus digitaler Öffentlichkeit /digitalen Lebenswelten

1 Baer & Sackofsky weisen in diesem Zusammenhang darauf hin, dass in der klassischen Rechtsauffassung Freiheit und Gleichheit als Gegenpole gefasst werden und Gleichheit hinter Freiheit zurücktritt, was aus feministischer Perspektive allerdings mit den skizzierten Problematiken einhergeht (Baer & Sacksofsky 2018, S. 14).

- Ungleichstellung zwischen den Geschlechtern in digitalen (Erwerbsarbeits-)Welten
- Berücksichtigung digitaler Gewalterfahrung und Reflexion digitaler Gewalt als gesellschaftliches Struktur- und Herrschaftsmuster
- Herstellung gleichberechtigter Autonomie in digitalen Lebenswelten

Gleichwohl stellt sich die Frage, wie sich diese Übersetzungen konkret darstellen: Was bedeutet Autonomie geschlechtertheoretisch gerahmt für den Prozess der Digitalisierung der Lebenswelt? Im Fokus meines Beitrages steht die Skizzierung von bzw. Sensibilisierung für Macht- und Geschlechterverhältnissen in Digitalisierungskontexten und -prozessen.

Ungleichberechtigte Autonomie-Verhältnisse in einer digitalisierten Lebenswelt

Die zuvor auf die Digitalisierung bezogenen Stränge der Kritik im Zusammenhang mit einer kritischen Hinterfragung des Autonomie-Begriffs verbinden sich mit zwei primären Fragestellungen: 1) Wer hat Zugang und gestaltet die Digitalisierung bzw. den digitalen Transformationsprozess (Machtfrage), und 2) wer wird in diesem Prozess ausgegrenzt bzw. ausgeblendet? Mit Blick auf das besonders im Fokus der Digitalisierung stehende Thema der Arbeit spitzt sich die Frage dahingehend zu, in welchen Bereichen und Branchen sich die Digitalisierung wie gestaltet bzw. gerade nicht und was dies für die Zukunft der Geschlechterverhältnisse bedeutet. Daran knüpft wiederum die Frage an, ob Digitalisierungsprozesse in Wirtschaft und in Gesellschaft eher als ein Risiko oder als eine Chance für die Geschlechtergerechtigkeit eingestuft werden müssen. So kommt Edelgard Kutzner für den Bereich der Erwerbsarbeit zu der Einschätzung, dass die Antwort auf diese Frage weniger von der Technik als vielmehr von der Organisation der Arbeit, der Arbeitsteilung und Arbeitsbewertung und damit der jeweils vorherrschenden Arbeitskultur abhänge (Kutzner 2020, zit. nach 3. GSB 2021, S. 26). Angesprochen sind hiermit zugleich die *Digitalisierungsfolgen mit ihren Effekten auf die Geschlechterverhältnisse* als zweite primäre Fragestellung.

Mit diesen beiden Fokussen, Gestaltungsmacht und Digitalisierungsfolgen, ist eine aus soziotechnischer Perspektive übergreifende Frage danach verbunden, wie die Digitalisierung als Technologie selbst als vergeschlechtlicht gedacht werden kann. Oder anders formuliert: Wie kommen Macht und Geschlecht in das Artefakt „Digitalisierung“ bzw. in die Digitalisierungsergebnisse? Wie konfigurieren sich einerseits Macht und Geschlecht innerhalb des Netzwerkes aus Entwickler*innen bzw. Produzent*innen, Nutzer*innen und andererseits Digitalisierungs-Artefakte? Diese Frageperspektive durchzieht die

folgende Analyse ebenfalls, wenn sich der Fokus auf die Gestaltungsmacht und die Digitalisierungsfolgen richtet. Vertiefend wird dieser Frage zudem mit Bezug auf das allgegenwärtige Digitalisierungsthema „Big Data“ im Zuschnitt der These von Big Data als „aufstrebende[s] neue[s] Wahrheitsregimes“ (Prietl 2019, S. 15) nachgegangen und dieses ebenfalls aus macht- und geschlechterkritischer Perspektive reflektiert.

Geschlechtersegregierte Gestaltungsmacht der Digitalisierung

Mit Blick auf den ersten Digitalisierungs-Kontext, d. h. dem ungleichen Zugang und der ungleichen Teilhabe an digitalen Lebenswelten, ist die Frage adressiert: *„Wer gestaltet die Digitalisierung?“* Allein schon mit dem Blick auf die Informatik als grundlegender Disziplin für die Digitalisierung und die Geschlechterverteilung unter den Absolvent*innen als einer demgemäß maßgeblichen Berufsgruppe der Digitalbranche wird deutlich, dass die Digitalisierung (nicht nur) in Deutschland auf einen besonders stark geschlechtstypisch segregierten Arbeitsmarkt trifft. So liegt der Frauenanteil in der Informatik im Jahr 2019 bei rund 22 % und ist über die letzten zwei Jahrzehnte um nur rund 7 % angestiegen (Destatis 2020, zit. nach Kompetenzzentrum Technik-Diversity-Chancengleichheit e. V 2021). In Führungspositionen sind Frauen gemessen an ihrem Anteil an allen Beschäftigten in der Digitalbranche zusätzlich noch einmal stark unterrepräsentiert (Hohendanner 2020, S. 37, zit. nach 3. GSB 2021, S. 44). Gleichheit als gleichberechtigte Autonomie zwischen den Geschlechtern hinsichtlich des Zugangs zu Gestaltungsmacht ist somit nicht gegeben, insbesondere, wenn man sich die Gründe für die weiterhin stabilen geschlechtssegregierten Berufswahl- und Karrieremuster vergegenwärtigt, die neben weiteren Gründen vor allem in starren Geschlechterstereotypen verortet werden (Jeanrenaud 2020, S. 30, zit. nach 3. GSB 2021, S. 43).

In der Folge sind Frauen bei der Entwicklung und Gestaltung von Technologie, aber auch bei der Entscheidung über ihren Einsatz stark unterrepräsentiert. Die Fragen, in welchen Bereichen digitale Technologien zum Einsatz und ob all ihre Möglichkeiten tatsächlich zum Einsatz kommen sollen, aber auch wer die politischen Richtungsentscheidungen vorgibt und wer über Zugänge zu relevanten Ressourcen und darüber entscheidet, welche Technologien von wem wie genutzt werden können, und ihre Antworten darauf, wirken sich auf die Geschlechterverhältnisse aus (3. GSB 2021, S. 18).

Die D21-Studie „Digital Gender Gap“ macht die ungleichen Gestaltungsmöglichkeiten unserer Lebenswelt im Kontext der Digitalisierung außerhalb professionalisierter Spezialbereiche sichtbar. Der D21-Digital-Index drückt als Durchschnittswert aus, wie gut die Gesellschaft mit der Entwicklung der Digitalisierung Schritt hält und liegt derzeit im mittleren Bereich. Der Indexwert wird aus den vier Subindices Zugang, Nutzungsverhalten, Kompetenz und

Offenheit gebildet. In jeder der vier Teilindices zeigen sich Unterschiede zwischen Frauen und Männern[2] unterschiedlichen Ausmaßes (Initiative D21 e. V. & Kompetenzzentrum Technik-Diversity-Chancengleichheit e. V. 2020, S. 6). Da dieser Index auf ungleiche Zugänge zu digitalen Informationstechnologien hinweist bzw. die unterschiedlichen Chancen der Nutzung von digitalen Informationstechnologien in der Lebenswelt anzeigt, liegt auch hier der Bezug zu einer fehlenden gleichberechtigten Autonomie auf der Hand, insofern über den Zugang zur Digitalisierung die möglichen Freiheitsgrade im beruflichen wie privaten Alltag maßgeblich beeinflusst werden. Die hieraus resultierende soziale Ungleichheit wird demgemäß mit den Begriffen „digital divide" und „digital inequality" (DiMaggio et al. 2004) beschrieben.

Intersektional aufbereitete Detail-Ergebnisse der Studie verdeutlichen zudem, dass es nicht „die" Frauen und „die" Männer gibt, sondern dass sich innerhalb der Geschlechtergruppen Unterschiede nach Alter, Bildung oder Tätigkeitsbereichen weiter ausdifferenzieren und damit auch zielgruppenspezifische Maßnahmen erforderlich machen, um den „digital divide" bzw. die digitale Ungleichheit abzubauen. Die vorfindbaren Geschlechterunterschiede in den Ergebnissen zu den vorhandenen digitalen Kompetenzen wie auch der Erweiterung des Wissenserwerbs lassen sich unter anderem auf Unterschiede in den mehr stark oder weniger stark durch Technik und Digitalisierung durchdrungenen Bildungs- und Berufsfeldern zurückführen, in denen Frauen und Männer tätig sind. So erfolgt zum Beispiel bislang kaum eine Vermittlung digitaler Grundkompetenzen in den weiblich geprägten Berufsfeldern (Initiative D21 e. V. & Kompetenzzentrum Technik-Diversity-Chancengleichheit e. V. 2020: 16), die wie zum Beispiel der Pflegebereich zudem als in der Regel digitalisierungsresistent eingestuft werden (s. u.).

Die Geschlechterunterschiede stellen sich darüber hinaus infolge von Geschlechterhierarchien am Arbeitsplatz ein und finden ihren Ausdruck unter anderem darin, dass nur bestimmten Gruppen die Möglichkeit zum mobilen Arbeiten eingeräumt wird. Hier zeigt sich wiederum am Beispiel von Büroarbeitsplätzen in der D21-Studie, dass vor der COVID19-Pandemie in Vollzeit berufstätige Männer gegenüber Frauen deutlich besser mit mobilen Geräten und digitalen Anwendungen ausgestattet waren. Flexibles Arbeiten, das ein selbstbestimmtes Arbeiten der Beschäftigten orts- und zeitunabhängig unterstützte, war damit bislang eher Männern als Frauen in der Arbeitswelt zugänglich (vgl. auch Carstensen 2019, S. 74, Kesselring & Vogl 2010). Es stellt sich nunmehr die Frage, ob bzw. wie umfassend sich alte Realitäten wiedereinstellen und inwieweit sich mit der Ermöglichung von Homeoffice nicht auch alte Herausforderungen der Vereinbarkeitsproblematik neu stellen. Der Aspekt des mobilen Arbeitens transformiert damit von der Machtfrage zur Digitalisierungsfolge.

2 Die Geschlechtskategorie „Divers" wurde in der Erhebung nicht berücksichtigt.

Geschlechterdisparate Digitalisierungsfolgen

Digitalisierung im Bereich der Erwerbs- und Fürsorgearbeit

Mit dem Blick auf die Zusammenhänge von Digitalisierung der (Erwerbs-) Arbeit und Geschlecht steht die Frage im Vordergrund, wie sich die Geschlechterverhältnisse durch Digitalisierung im Bereich der Erwerbs- und Sorgearbeit zukünftig gestalten könnten. Grundsätzlich bestehen weiterhin Entwicklungspotenziale hinsichtlich der Erhöhung des Anteils (voll-)erwerbstätiger Frauen und einer geschlechtergerechteren Gestaltung der Arbeitsteilung in der Sorgearbeit. Auch die Verbesserung der Vereinbarkeit von Beruf und Familie – nicht nur für Frauen – ist weiterhin ein zentrales Thema, um die Gleichberechtigung der Geschlechter in der Erwerbsarbeitswelt und im Privaten einzulösen (Oliveira 2017, S. 70 ff.) und damit dem Anspruch einer gleichberechtigten Autonomie gerecht zu werden.

Mit der Digitalisierung von Erwerbsarbeit bzw. Arbeit 4.0, zumeist aber als Industrie 4.0 im Zentrum, wird zunehmend relevant, wie sich nicht nur erneut die Produktions- bzw. Arbeitsverhältnisse verändern, sondern damit auch die Geschlechterverhältnisse. Im Kern sind es vor allem vier mit einander verbundene Phänomene, die die vierte Revolution in der Erwerbsarbeit kennzeichnen: die exponentielle Leistungssteigerung im Bereich der Informations- und Kommunikationstechnologie und in Folge immer leistungsfähigere Rechner, die Vernetzung von Menschen und Dingen im Internet (Stichwort „Internet of Things“), neue bzw. weiterentwickelte Technologien im Bereich der Robotik und Sensorik, aber auch Smartphones und der 3-D-Druck, sowie die Sammlung und Auswertung großer Datenmengen (Stichwort „Big Data“) (BMAS 2016, zit. nach Schildmann 2018, S. 4). Diese Phänomene erzeugen eine erhebliche Veränderungsdynamik in vielen Bereichen der Gesellschaft und in der Alltagswelt. Branchengrenzen verschieben sich, es kommt zu Veränderungen in den Wertschöpfungsketten, der Arbeitsorganisation, in den Tätigkeitszuschnitten sowie den Einkommensmöglichkeiten (ebd.). Der Wandel betrifft somit fast alle Aspekte des Arbeitsmarktes. Unbestritten ist jenseits einer überzogenen Glorifizierung bzw. Verteufelung der Digitalisierung, dass dieser Prozess die Arbeitswelt von Frauen und Männern gegenwärtig und auch zukünftig verändern wird. Die Qualität und die Richtung des Wandels bleibt indes noch unbestimmt (Stettes 2016) bzw. wird insbesondere für die hochqualifizierte Wissensarbeit aus geschlechterpolitischer Perspektive eher als ambivalent eingeschätzt (Will-Zocholl & Clasen 2020, 100). Vor diesem Hintergrund finde sich eine Reihe von Studien in der Geschlechterforschung, die nach den Möglichkeitsräumen fragen, aber auch welche Risiken für die Gestaltung einer geschlechtergerechten Arbeitswelt sich öffnen (Marrs et al. 2015, Lott 2015, Schildmann 2018).

Christina Schildmann (2018) führt diesbezüglich fünf Felder auf, auf denen sich das Geschlechterverhältnis auf dem Arbeitsmarkt zukünftig entscheiden könnte: Mobiles Arbeiten als ermöglichender Faktor räumlicher und zeitlicher Flexibilisierung von Arbeit (s. o.) und insbesondere, so ist an dieser Stelle zu ergänzen, im Bereich der hochqualifizierten Wissensarbeit (vgl. Will-Zocholl und Clasen 2020), Digitalisierung des Dienstleistungsbereiches, soziale Dienstleistungen als digitalisierungsresistente Seite der Ökonomie, digitale Plattformen und der Start-up-Bereich. Diese Bereiche möchte ich nachfolgend etwas eingehender betrachten, insofern sie auch aus einer geschlechterkritischen Autonomie-Perspektive im besonderen Maße relevant sind.

Hochqualifizierte Wissensarbeit

Vor allem im Kontext hochqualifizierter Wissensarbeit ist auch bereits vor der Corona-Pandemie eine Entgrenzung von Arbeit hinsichtlich zeitlicher und räumlicher Flexibilisierung sowie eine Zunahme prekärer Beschäftigungsbedingungen im Zug von Globalisierungs- und Digitalisierungsprozessen zunehmende Realität (vgl. Beiträge in Flecker 2016, Will-Zocholl & Clasen 2020, S. 100). Demgemäß kommen Will-Zocholl und Clasen (2020, S. 118) zu der Einschätzung, dass die Digitalisierung mit ihren Möglichkeiten einer flexibleren Arbeitsgestaltung, insbesondere im Hinblick auf egalitäre Geschlechterverhältnisse, ambivalent und eher negativ zu werten sind. In ihrer Studie über Ingenieurinnen in der Industrie als ein Beispiel für hochqualifizierte Wissensarbeit zeigen Will-Zocholl und Clasen (2020) auf, dass örtliche Flexibilisierung durch Digitalisierung und Virtualisierung eine Zunahme physischer Mobilität mit sich bringt, die zeitliche Flexibilisierung das Risiko eines dauerhaften Verfügbarkeitsanspruches in sich birgt und insbesondere in international agierenden Unternehmen die Unsicherheit durch die Konkurrenz in globalem Maßstab erhöhen kann. Entgrenzung, Flexibilisierung und Prekarisierung stehen damit unmittelbar im Zusammenhang mit der Autonomie von hochqualifizierten Beschäftigten und dies in ambivalenter Art und Weise. Auch mobiles Arbeiten und Homeoffice wurden bereits vor der Corona-Krise nicht einmütig als Faktoren, wenn nicht gar als Heilsbringer gefeiert, die zu einer besseren Vereinbarkeit von Familie und Beruf für Frauen und Männer beitragen würden. Einerseits wurde argumentiert, dass mobiles Arbeiten und Homeoffice helfen könne, „die Schere zwischen den Arbeitszeiten von Frauen und Männern zu schließen und die Präsenzkultur in den Betrieben aufzubrechen. Gleichzeitig bergen sie [mobiles Arbeiten und Homeoffice aber auch] die Gefahr, dass Beruf und Arbeit untrennbar verschmelzen, dass die Entgrenzung von Arbeit die Vereinbarkeit weiter erschwert und die Unsichtbarkeit im Homeoffice sich negativ auf die Berufslaufbahn auswirkt“ (Schildmann 2018, S. 5, vgl. auch Arntz et al. 2019). Zudem wurde befürchtet, dass Homeoffice „für Männer und Frauen unterschiedliche Effekte erzeugt: So besteht die Gefahr, dass Mütter das

Homeoffice in erster Linie für Vereinbarkeitsanliegen nutzen; Väter hingegen ihren ohnehin langen Arbeitstag verlängern, um Karrierepunkte zu sammeln – oder dies zumindest in den Köpfen der Vorgesetzten stattfindet“ (Schildmann 2018, S. 5). Unter den Bedingungen von Home Schooling scheint sich diese Gefahr nicht nur bestätigt, sondern auch verschärft zu haben. Detailanalysen stehen hierzu noch aus. Vor der Krise war, wie gezeigt, mobiles Arbeiten allerdings ohnehin nur einer Minderheit der Arbeitnehmer*innen möglich gewesen. Wie sich die Möglichkeiten zum Homeoffice nach der Krise entwickeln werden, ob die Potenziale von Homeoffice gehoben werden können, wird sich also erst noch zeigen. In flexibilitätsoffenen Branchen wie der des Ingenieurwesens deutete sich allerdings an, dass insbesondere unter Zunahme von Unsicherheit Präsenzzeiten nach wie vor den Vorzug gegeben wird (Will-Zocholl & Clasen 2020, S. 117).

Somit kommen auch aktuelle Studien weiterhin zu dem Ergebnis einer Retraditionalisierung im Geschlechterarrangement (Krüger 2009, S. 71 f.), spätestens dann, wenn Kinder im Haushalt sind (Panova et al. 2017).

Digitalisierung des Dienstleistungsbereiches und Digitalisierungsresistenz sozialer Dienstleistungen

Ist die Digitalisierung der Erwerbsarbeit Thema, so liegt der Fokus zumeist auf der industriellen Produktion und deren Automatisierung, die überwiegend männlich besetzt ist. 2020 hat das BMBF das Positionspapier der Plattform Industrie 4.0 vorgestellt, aus dem hervorgeht, wie Industrie 4.0-Lösungen zum Beispiel einen Beitrag zur Bewältigung der wirtschaftlichen Folgen von COVID-19 leisten können. Aber auch in den produktionsfernen Bereichen der Industrie und in den Dienstleistungsbranchen floriert die Digitalisierung. Das gilt für die unternehmensbezogenen Dienstleistungen genauso wie für Handel, Banken, Versicherungen und Logistik. Diese Branchen sind gegenüber der Industrie 4.0 stärker weiblich dominiert. So haben Andreas Boes et al. dargelegt, dass die digitale Transformation durch umfassende Informationssysteme wie SAP zunehmend die Verwaltung rationalisieren und sich damit in die „indirekten“ Kopfarbeitsbereiche der Angestellten ausgebreitet (vgl. Boes et al. 2016). Durch die Digitalisierung werden alle Aktivitäten und Prozesse messbar wie auch transparent. Hier wächst ein Spannungsfeld zwischen Autonomie und Kontrolle, das gestaltet werden kann bzw. muss: Lean Management und Digitalisierung werden damit zum Handlungsfeld von Betriebsräten und Gewerkschaften. Allerdings sind viele weiblich dominierte Arbeitsbereiche nicht gewerkschaftlich organisiert und entziehen sich damit einer Interessenvertretung der Arbeitnehmenden. Die Digitalisierung im Dienstleistungssektor bringt damit ebenfalls große Dynamiken hervor, die auf der einen Seite mit Jobverlusten, auf der anderen Seite mit Veränderungen von Tätigkeiten und

in der Folge Weiterbildungs- und Umschulungsbedarf einhergehen (Stettes 2016, S. 30).

Der Industrie 4.0 und den zuvor genannten Dienstleistungsbranchen steht mit den sozialen Dienstleistungen und mit 80 bis 90 Prozent Frauenanteil eine Branche gegenüber, die zukünftig vor dem Hintergrund des demografischen Wandels und einem fortschreitenden Anstieg der Erwerbstätigkeitsrate von Frauen stark an Bedeutung zunehmen wird. Diese Branche ist zudem gleichzeitig am stärksten „digitalisierungsresistent“ (Baethge & Baethge-Kinsky 2017). Es gehen also nicht nur die geburtenstarken Jahrgänge auf das Rentenalter zu, was zu einem steigenden Beschäftigtenbedarf bei den Pflegedienstleistungen führt; sondern mit der höheren Erwerbstätigkeit von Frauen benötigen immer mehr berufstätige Eltern eine Betreuung für ihre Kinder. Vor dem Hintergrund dieser Entwicklungen sind die Debatten über die Aufwertung der sozialen Dienstleistungen aufgrund des teils schon realen, teils drohenden Fachkräftemangels (Flake et al. 2018) dennoch eher verhalten. Auch in der Diskussion über die Arbeit der Zukunft kommt dieser Branche noch eine nachrangige Bedeutung zu. Während sich die Debatte technikzentriert darauf richtet, dass Deutschland im Wettlauf um die schnellste Digitalisierung abgehängt werden könnte, könnte sich vielmehr, darauf weist Schildmann (2018, S. 6) hin, die finanzielle Unterausstattung der sozialen Dienstleistungen ebenfalls als hoch relevante Schwachstelle der deutschen Ökonomie erweisen. Dass sich aber auch in dieser Branche Digitalisierungsprozesse bereits eingestellt haben, die ebenfalls die Autonomie der in der Pflege Beschäftigten in ambivalenter Art und Weise beeinflussen, zeigt Lena Weber (2020) in ihrer Studie über neue Formen der Rationalisierung in der Pflege durch den Einsatz digitaler Technologien auf.

Plattformökonomien und digitale Start-Ups

Auf digitalen Plattformen[3] erwächst zunehmend ein neuer Arbeitsmarkt, auf dem sich neue Jobs, insbesondere auch für Frauen, ergeben. Dieser ist gekennzeichnet durch vornehmlich ungesicherte Arbeitsverhältnisse, die jenseits von betrieblicher Mitbestimmung, gewerkschaftlicher Vertretung und arbeitsrechtlicher Regulierung verortet sind (Benner 2015, zit. nach Schildmann 2018, S. 6) und neue Räume für Gewalt ermöglichen (Frey 2020, S. 18). Derzeit ist der Einfluss von Crowdwork in Deutschland mit ca. drei Millionen Personen noch gering. Zumeist handelt es sich bei den Personen um solche, die die Crowdwork als Zweitjob ausüben (Stettes 2016, S. 44). Es gibt aber auch bereits Anzeichen dafür, dass sich hier ein weiterer „Frauenarbeitsmarkt“ entwickelt, insbesondere wenn die Tätigkeiten im Betreuungs-, Pflege- und Rei-

3 Digitale Plattformen ermöglichen das Outsourcing betrieblicher Tätigkeiten in eine Cloud, die dort von einer anonymen „Crowd“ abgerufen und verrichtet werden. Dieses Phänomen wird als „Plattformisierung“ bzw. „Crowdsourcing“ bezeichnet.

nigungssektor verortet sind. So bieten die Mikrojobs eine niedrigschwellige Vereinbarkeitslösung für Sorgeverantwortliche, in der Regel Mütter, da diese zeitlich und z. T. räumlich flexibel erledigt werden können. Solche Plattformen bieten legal und niedrigschwellig kostengünstige soziale und haushaltsnahe Dienstleistungen an, mit denen wiederum die Vereinbarkeitsprobleme von in der Regel hoch qualifizierten privaten Auftraggeber*innen gelöst werden können. Der Bedarf an diesen Dienstleistungen wächst kontinuierlich und birgt das Risiko eines wachsenden Bereiches von prekärer Soloselbstständigkeit in sich. Die Herausforderung besteht somit darin, die Vermittlung von sozialen und haushaltsnahen Dienstleitungen über Internetplattformen derart zu gestalten, dass „gute" Arbeit entsteht (Schildmann 2018, S. 8, vgl. auch Meier-Gräwe 2016).

Demgegenüber ist die Start-up-Szene, die zum Sinnbild der „digitalen Disruption" (Schildmann 2018, S. 8) und der modernen Wirtschaft geworden ist, mit über 80 % wiederum stark männlich dominiert (ebd.), und vermutlich auch weiß, jung und gut gebildet. Hier zeigen sich mit Blick auf die Vergabe des Risikokapitals insbesondere für Frauen geringere Chancen. Zudem sind Sexismus und „unequal pay gaps" offensichtlich ein verbreitetes Phänomen (The Economist 2017). Da es sich bei den Start-ups potenziell um einflussreiche Unternehmen der Zukunft handelt, besteht insbesondere in diesem Gründungsbereich Handlungsbedarf hinsichtlich des Abbaus von Geschlechterstereotypen und der finanziellen Förderung von Frauen.

Digitalisierungsfolgen: Sexismus und Gewalt

Gegenüber den noch eher unbestimmten Auswirkungen der Digitalisierung auf die Arbeitsplatzentwicklung in den jeweiligen Branchen, stellen Sexismus und Gewalt im Internet dagegen eine Digitalisierungsfolge dar, die „analoge" Ausdrucksformen und Praktiken auf digitale Formate und Öffentlichkeit transferiert und damit eine Fortsetzung oder Ergänzung von Gewaltverhältnissen und -dynamiken darstellt. Zugleich kommt digitaler Gewalt eine neue Qualität zu, da digitale Instrumente neue Formen einer geschlechtsbezogenen Gewaltausübung ermöglichen. Zudem erschweren fehlende Raum- und Zeitbarrieren sowie Anonymität und Identitätsdiebstahl die Bekämpfung digitaler Gewalt (3. GSB 2021, S. 125). Die dargelegten Beispiele im 3. GSB zeugen von der enormen Bandbreite sexualisierter und geschlechtsbezogener digitaler Gewalt bzw. mittels digitaler Technik, die Verwirklichungschancen und damit die Autonomie insbesondere für Frauen[4] massiv einschränken, sowohl im privaten Leben und im Erwerbsleben, als auch hinsichtlich der Teilhabe an demokratischer Willensbildung und -äußerung (Frey 2020, S. 1). Allerdings

4 Männer können ebenfalls von geschlechtsbezogener (digitaler) Gewalt betroffen sein, wenn sie zum Beispiel bestimmten stereotypen Männlichkeitsvorstellungen nicht entsprechen oder sich emanzipativ äußern (Frey 2020, S. 5).

kann digitale Technik auch zum Schutz vor (digitaler) Gewalt eingesetzt werden. Ein Beispiel ist die unter dem Hashtag #MeToo ausgelöste weltweite Debatte über Sexismus und sexualisierte Gewalt gegen Frauen (Hasenkamp 2020).

Digitalisierungsfolgen: Geschlechterungleichheit durch Big Data, Algorithmen und Künstliche Intelligenz (KI)

Die Machtfrage stellt sich nicht nur in gesellschaftlichen Produktions- und Reproduktionsverhältnissen mit ihren Auswirkungen auf individuelle Autonomiespielräume, sondern auch bei der Interpretation digital verfügbarer Daten. In einer zunehmend datafizierten Gesellschaft und mit der bereits starken Durchdringung des Alltags mit Algorithmen verteilen sich die Chancen von Autonomiegewinnen bzw. Kontrollrisiken von „Big Data" in komplexer und zum Teil widersprüchlicher Form. So zum Beispiel, wenn es darum geht, wer was über wen weiß, und dieses Wissen zudem den Wissenslieferant*innen verborgen bleibt. Big Data und deren algorithmische Verarbeitung stellt damit ein Instrument dar, das unterschiedliche Interessen befördern und verschiedene Machtverhältnisse unterstützen kann, insbesondere, wenn Werte und Privilegien, die durch die kodierten Regeln festgelegt werden, in Black Boxes verborgen bleiben (Pasquale 2015, vgl. auch 3. GSB 2021, S. 35).

Bianca Prietl (2019) analysiert aus einer geschlechterkritisch informierten, wissenschaftstheoretischen Perspektive drei große Versprechen von Big Data bzw. deren damit verbundene erkenntnistheoretische Prämissen, wie sie in den Diskursen über Big Data aufscheinen (vgl. Prietl 2019, S. 15). Diese fasst sie mit den Begriffen Datenfundamentalismus, post-explanativer Antizipationspragmatismus und anti-politischer Solutionismus. Ihre Analysen verdeutlichen, dass die Big Data zugrundeliegenden Prämissen ungleiche Autonomiespielräume gesellschaftlicher Gruppen affirmieren. Ihre Analysen können damit zugleich als eine gleichstellungsorientierte (Technik-)Folgenabschätzung (Hummel et al. 2020: 10) verstanden werden, die jenseits eines stereotyp-binär verfassten Genderverständnisses und materialistischen Technikverständnisses ansetzt (Both 2017, zit. nach Hummel et al. 2020, S. 60).

Datenfundamentalismus als neue Erkenntnisprämisse bezieht sich auf das Verfahren des Maschinellen Lernens, bei dem selbstlernende Algorithmen selbstständig in großen Datensätzen nach Mustern suchen und derart auch unvorhersehbare Zusammenhänge aufdecken. Wenn aber in diesem Vorgehen der Vorstellung Vorschub geleistet wird, dass Korrelation auf Kausalität hinweist und dass massive Datensätze und prädiktive Analysen und damit Prognosen, immer oder zunehmend objektive Wahrheit widerspiegeln, dann wird ein positivistischer, datafizierender Weltzugriff deutungsmächtig. Dieser blendet solche Aspekte der sozialen Wirklichkeit aus, die sich einer einfachen datenförmigen Repräsentation entziehen. Mit dem Datenfundamentalismus

besteht somit nach Prietl (2019) auch das Risiko, dass die tradierte Vorstellung westlicher, männlich konnotierter Rationalität und Objektivität erneut deutungsmächtig werden und ein Einfallstor für geschlechterbezogene Ausschlüsse eröffnen. An den Datenfundamentalismus schließt ein post-explanativer Antizipationspragmatismus an, d.h. das Treffen probabilistischer Vorhersagen aus Big Data-Analysen. Damit geht die Gefahr unzulässiger bzw. fehlerhafter Schlussfolgerungen aus diesen Big Data-Analysen einher. Ein viel zitiertes Beispiel in dieser Hinsicht ist ein von Amazon zur Automatisierung von Personalrekrutierungen entwickelter Algorithmus, der im Ergebnis systematisch die Bewerbungen von Männern bevorzugt hat. Mittlerweile werden Algorithmen zur Automatisierung von Personalrekrutierungen bereits in multinationalen Konzernen wie zum Beispiel Ernest & Young eingesetzt (Elsherbiny 2020), unklar bleibt inwieweit in den Konzernen mit der bekannten Problematik der Bias-Verstärkung durch Algorithmen umgegangen wird.

Datenfundamentalismus und post-explanativer Antizipationspragmatismus sind reduziert auf die Bearbeitung von Was- oder Wer-Fragen. Somit geht es nicht mehr darum, Phänomene zu verstehen und zu erklären und damit um Warum- oder Wie-Fragen. Diese Erkenntnismodelle erlauben Vorhersagen darüber, wie sich betrachtete Variablen in Zukunft gemeinsam entwickeln werden. Dieses geschieht allerdings auf der Grundlage von Daten, die aus der Vergangenheit stammen und die Vorhersagen ohne Kausalitäten begründen. Weitere prominente Beispiele für das Fruchtbarmachen probabilistisch-prädiktativen Wissens für die antizipierende Gestaltung und Regierung der Zukunft finden sich in der Verbrechensbekämpfung, dem Strafvollzug oder der Kreditvergabe, die ebenfalls jeweils nicht geschlechtsneutral sind. Die Big Data zugrundeliegende „post-explanative Pragmatik“ (Andrejevic 2014, S. 1675, zit. nach Prietl 2019, S. 19) birgt in Kombination mit dem Antizipationsregime die Gefahr, soziale Ungleichheits- und Herrschaftsverhältnisse zu affirmieren und zu reproduzieren. Die erhoffte Unabhängigkeit von menschlichen Einflüssen verhilft nicht zu mehr Neutralität, sondern droht zur unausweichlichen technologischen Fortschreibung diskriminierender sozialer Strukturen zu werden.

Die inzwischen vielfach adressierte (Re)Produktion von sozialen Ungleichheiten qua datenbasierter Positions- und Ressourcenzuweisung (Mau 2017) basieren letztendlich auf Entscheidungen, die von fehlbaren Menschen getroffen werden. Als menschliche Vorurteile, Missverständnisse und Voreingenommenheiten fließen sie als Verzerrungseffekte (Bias) in die Algorithmen und Softwaresysteme ein, die in immer höherem Maße unseren Alltag bestimmen. Die unbewussten Einschreibungen von Geschlechterannahmen können sich aber in Big Data-Analysen wie auch in andere technische Produkte grundsätzlich einschleichen, wenn keine explizite Zielgruppe im Fokus steht und eine Technik- bzw. Produktentwicklung für „jedermann“ erfolgt (Rommes et al. 2012, van Oost 2003). Eine derartige Entwicklungspraxis läuft grundsätzlich Gefahr, gesellschaftliche Ungleichheitsstrukturen aufrecht zu erhalten

sowie hegemonialen Maskulinitätsvorstellungen und ihren Normen zu folgen. Die Männerdominanz in technischen Gestaltungskontexten verstärkt diese Prozesse und befördert eine maskulin geprägte Technikkultur.

Hieran knüpft die dritte von Prietl (2019) analysierte Prämisse bzw. das damit verbundene Versprechen eines „anti-politischen Solutionismus“ an. Protagonisten (selten Protagonistinnen) der New Economy und die Befürworter von Big Data imaginieren sich als Heilsbringer, die die Fehler und Probleme der Menschheit mittels neuester Informations- und Datentechnologien lösen. Mit der Fokussierung von Big Data auf ausschließlich sachbezogene Probleme und deren technische Lösungen erweisen sich diese Lösungen zugleich wenig sensitiv für Machtasymmetrien, Interessenkonflikte und Ungleichheitsstrukturen. Als Quellen menschlicher Potenzialentfaltung gelten für Solutionisten die Vernetzung von Informationen, Menschen und Dingen. Gefordert wird vor diesem Hintergrund zwar die Gleichheit im Zugang zu Informationen, Wissen und Technik sowie ein Imperativ der Meritokratie. Ausgeblendet werden dabei aber die Ungleichheit im Zugang zu digitalen Technologien und sich im virtuellen Raum reproduzierende Machtsymmetrien sowie die ungleichheitsstabilisierenden Effekte eines unkritisch bekräftigten Meritokratie-Ideals. Mit Big Data werden zudem Tätigkeitsfelder an Bedeutung gewinnen, die aufgrund der eingangs dargelegten stark geschlechtssegregierten Berufs- und Studienfachwahl von Männern dominiert werden. Im anti-politischen Solutionismus artikuliert sich damit nicht zuletzt die tendenzielle Machtvergessenheit von Akteur*innen, „die ihrer eigenen Position als Herrschende nicht gewahr sind“ (Prietl 2019, S. 21).

Offensichtlich wird an den Analysen von Prietl (2019) hinsichtlich der zentralen Prämissen und Versprechen von Big Data bzw. allgemein an der digitalen Transformation der Gesellschaft, dass diese einer geschlechterkritisch gefassten Autonomie-Perspektive in vielen gesellschaftlichen Feldern zuwiderlaufen. Die Prämissen und Versprechen unterlaufen einen gleichberechtigten Zugang zu Ressourcen und zur Teilhabe an der Digitalisierung der Lebenswelt. Hier besteht demnach dringender Handlungsbedarf, um soziale Ungleichheit zukünftig nicht zu verstärken. Hier setzt der 3. GSB an. Sein Ziel war es, gleichstellungsorientierte Verwirklichungschancen im Kontext der digitalen Transformation zu bewerten und Handlungsempfehlungen zur Förderung einer solchen zu erarbeiten. Ein übergreifendes Instrument stellt dabei die gleichstellungsorientierte Gesetzes- und Technikfolgenabschätzung dar. In der im Auftrag des 3. GSB vorgelegten Analyse von Hummel et al. (2020) machen die Autor*innen die gleichstellungsorientierte Gesetzes- und Technikfolgenabschätzung für eine entsprechende Folgenabschätzung im Kontext der Digitalisierung fruchtbar. Hier zeigt sich, dass bislang für die Technikfolgenabschätzung im Allgemeinen und für die Digitalisierung im Besonderen kaum die Erkenntnisse feministischer Theorien und der Geschlechterforschung aufgegriffen und in die Analysen einbezogen werden. Entsprechend gehen die

Handlungsempfehlungen an die Politik dahin, dass gleichstellungsorientierte Perspektiven in die Gesetzes- und Technikfolgenabschätzung einerseits über die Institutionen selbst sowie andererseits durch Verfahren der Folgenabschätzung integriert werden könnten (3. GSB 2021, S. 155 f.). Dabei ist es wichtig, dass die Folgenabschätzung bereits zu einem frühen Zeitpunkt des Entwicklungsvorhabens erfolgt, um Geschlechterungleichstellung im Planungs-, Entwicklungs- und Gestaltungsprozess proaktiv zu vermeiden. Es sind vor allem die Entwickler*innen, die vielfältige Möglichkeiten haben, auf die Beziehung zwischen Nutzer*innen und Artefakten der Digitalisierung Einfluss zu nehmen. Entwickler*innen treffen zum Beispiel die Auswahl über Funktionen, nehmen über eine Hierarchisierung eine Bewertung von Bedeutungen vor oder verteilen Aufgaben und Handlung zwischen Mensch und Artefakt bzw. Nutzendem und System (Buchmüller 2018, S. 64). Die Gestaltungsresultate sind somit Ergebnisse verschiedenster Erwägungen und Entscheidungen, in die auch geschlechtliche Vorstellungen – mehr oder weniger bewusst – mit einfließen. Aber Menschen können sich auch an algorithmische Ergebnisse anpassen. Eine solche Anpassung enthält normative Vorstellungen, wie die Ergebnisse der Algorithmen als (a-)normal interpretiert werden. Insofern kommt nicht nur den Technikentwickler*innen, sondern auch der (Non-)User-Integration in die Verfahren der Produkt- und Technikentwicklung eine herausragende Bedeutung zu (Augustin et al. 2021).

Fazit und Perspektive der Schaffung einer gleichberechtigten Autonomie in digitalen Lebenswelten

Die Datafizierung und digitale Transformation der Gesellschaft bringt ohne eine geschlechterkritisch informierte Reflexion weiterführende soziale Ungleichheit hervor, denn hierauf weist auch Jutta Weber hin: „Die technische Rekonfiguration der Welt [… ist] kein neutrales Fortschrittsprojekt, sondern eine umkämpfte und höchst politische Angelegenheit, die von vielen Akteur_innen mit unterschiedlichen Interessen, Perspektiven und Rhetoriken vorangetrieben wird“ (Weber 2017, S. 357, zit. nach Prietl 2019, S. 15). Eine kritische Analyse der Digitalisierungs- und Datafizierungsprozesse stellt insofern aus gesellschaftspolitischer und damit zugleich geschlechterkritischer Perspektive eine Notwendigkeit dar. Hieran anknüpfend war es ein Anliegen dieses Beitrages, unter dem Fokus eines geschlechterkritischen Verständnisses von Autonomie die ungleichheitsrelevanten Dimensionen der digitalen Transformation im Kontext bestehender Geschlechterverhältnisse zu untersuchen. Hierfür wurden vier zentrale geschlechterbezogene Autonomieperspektiven auf die Digitalisierung angewendet und auf zwei leitende Fragestellungen

hin analysiert, die ihren Fokus auf Machtverhältnisse und die Folgen in Digitalisierungsprozessen richteten. Die aus diesem aufgespannten Analyserahmen hervorgehenden Erkenntnisse verdeutlichen, dass Digitalisierungsprozesse einem besonderen gesellschaftlichen Gestaltungsbedarf unterliegen, damit diese einer demokratischen, sozialen und geschlechtergerechten sowie vielfältigen Gesellschaft gerecht werden und damit nicht zuletzt die Autonomie der Gesellschaftsmitglieder einer digitalisierten Lebenswelt bzw. in einer datafizierten Gesellschaft gleichermaßen ermöglichen.

Wie für den (hoch-)qualifizierten Technologiesektor insgesamt gilt für die der Digitalisierung zugrundeliegenden Informatik- und Informationstechnikbranche in der Berufswelt wie auch im Alltag weiterhin ein geschlechtssegregierter Zuschnitt nach altbekanntem Muster, der Frauen tendenziell auf technikferner und Männer auf technikzentrierter Seite verortet. In der Folge stellen sich die Digitalisierungsfolgen für Frauen und Männer geschlechterdisparat dar – nicht zuletzt hinsichtlich des Anspruchs auf datenbezogene Autonomie. So zeigen sich geschlechterdisparate Prozesse der digitalen Transformation im Zugang, in der Nutzung und in der Gestaltung digitaler Technologien. In diesem Kontext sind Zeitsouveränität, Raumsouveränität und informationelle Selbstbestimmung relevant, deren Ungleichheit sich in (digitalen) Arbeitskulturen, unterschiedlichen Chancen und Risiken der Flexibilisierung, in der Zuschreibung von Kompetenzen oder auch in Unterschieden mit digitalen Gewalterfahrungen zeigen. Auch die Ergebnisse algorithmisch erzeugter Analysen erweisen sich infolge männlicher Überrepräsentanz in den genannten Branchen und gesellschaftlich materialisierter Stereotype mitunter geschlechtlich verzerrt. Dies manifestiert in der Folge Ungleichheitsverhältnisse.

Insbesondere Big Data und deren Verarbeitung durch Algorithmen und Künstliche Intelligenz stellen somit nicht nur eine Technik der Wissensproduktion dar, sondern zugleich auch eine Machtform. Hier wird die Verantwortung der verschiedenen Akteur*innen im soziotechnischen System „Digitalisierung“ für die Gewährleistung einer gleichberechtigten Autonomie im Zugang, in der Nutzung und Gestaltung in besonderem Maße deutlich. Die Macht der Algorithmen, die durch Expert*innen aus den Technikwissenschaften entwickelt werden, liegt in ihrer Fähigkeit, Entscheidungen zu treffen, zu klassifizieren, zu sortieren, zu ordnen. Mit dem sozialen Phänomen Big Data wird ein Objektivitätsverständnis erneut deutungsmächtig gemacht, das Erkenntnisweisen und Wirklichkeitsaspekte an westlich-weiße Männlichkeit knüpft und andere Erkenntnisweisen zu marginalisieren droht. Zudem erkennen große Teile der Gesellschaft und hierunter viele Frauen in der Digitalisierung kaum Vorteile, wenn digitale Innovationen und die digitale Transformation mit vorwiegend auf technische Optimierung und Vernetzung ausgerichtete Entwicklungen in Wirtschaft und Gesellschaft einhergehen. Digitale Technologieentwicklung muss stärker als bislang unterschiedliche Zugänge hinsichtlich Geschlechter-, Alters-, Bildungs- oder fachkultureller Aspekte

berücksichtigen. Diese Vielfalt an Zugängen muss auch in als wenig technikaffin betrachtete, berufliche Sektoren, wie im Bereich der Pflege und Gesundheit sowie dem privaten reproduktionsbezogenen Sektor, eingebracht werden (Initiative D21 e. V. & Kompetenzzentrum Technik-Diversity-Chancengleichheit e. V. 2020, S. 9).

Damit sich die aktuelle digitale Technologieentwicklung nicht auch weiterhin an männlichen Sichtweisen und Maßstäben orientiert und aus ungleichen Chancen der Nutzung und der Gestaltung eine weiterführende u. a. geschlechterbezogene Ungleichheit resultiert, muss digitale Technologieentwicklung, vornehmlich Big Data und KI, zukünftig stärker als bislang auch unter Berücksichtigung einer geschlechterkritischen Autonomieperspektive gestaltet werden. Im 3. GSB finden sich eine Vielzahl von Handlungsempfehlungen, die nunmehr der Umsetzung und Weiterentwicklung bedürfen.

Literatur

Arntz, M., Yahmed, S. B. & Berlingieri, F. (2019). *Working from Home: Heterogeneous Effects on Hours Worked and Wages.* ZEW Discussion Paper No. 19–015.

Augustin, L., Kokoschko, B., Wolffram, A. & Schabacker, M. (2021). Defining the Non-user: A Classification of Reasons for Non-use. In A. Chakrabarti, R. Poovaiah, P. Bokil & V. Kant (Hrsg.), *Design for Tomorrow.* Volume 1. Singapore: Springer, S. 339–349.

Baer, S. & Sacksofsky, U. (2018). Autonomie im Recht geschlechtertheoretisch vermessen. In Dies. (Hrsg.), *Autonomie im Recht. Geschlechtertheoretisch vermessen.* 1. Auflage. Nomos, S.11–30.

Baethge, M. & Baethge-Kinsky, V. (2017). *Entwicklung des Arbeitsmarktes unter geschlechtsspezifischen Aspekten mit einem Exkurs zu Frauenerwerbstätigkeit und Digitalisierung. Expertise im Rahmen des Zweiten Gleichstellungsberichts der Bundesregierung*, www.gleichstellungsbericht.de/kontext/controllers/document.php/25.e/0/83159e.pdf [Zugriff: 3. 5. 2021].

Boes, A., Kämpf, T., Langes, B. & Lühr, T. (2016). *„Lean" und „agil" im Büro. Neue Formen der Organisation von Kopfarbeit in der digitalen Transformation*, Working Paper Forschungsförderung, Nr. 23. Düsseldorf: Hans-Böckler-Stiftung, www.boeckler.de/pdf/p_fofoe_WP_023_2016.pdf [Zugriff: 03. 05. 2021].

Buchmüller, S. (2018). *Geschlecht Macht Gestaltung – Gestaltung Macht Geschlecht. Der Entwurf einer machtkritischen und geschlechterinformierten Designmethodologie.* Logos Verlag Berlin.

Carstensen, T. (2019). Verunsichtbarung von Geschlechterungleichheiten? Digitalisierte Arbeit zwischen Rhetoriken neuer Möglichkeiten und der Reorganisation alter Muster. In B. Kohlrausch, Bettina, C. Schildmann & D. Voss (Hrsg.), Neue *Arbeit – neue Ungleichheiten? Folgen der Digitalisierung. 1. Auflage. Arbeitsgesellschaft im Wandel.* Beltz Juventa, S. 69–87.

Carstensen, T. (2015). *Im WWW nichts Neues. Warum die Digitalisierung der Arbeit Geschlechterverhältnisse kaum berührt.* Luxemburg. Gesellschaftsanalyse und

linke Praxis 3/2015. https://www.zeitschrift-luxemburg.de/im-www-nichts-neues/ [Zugriff: 24.03.2021].

Connell, R. W. (1999). *Der gemachte Mann: Konstruktion und Krise von Männlichkeiten.* Leske + Budrich.

DiMaggio, P., Hargittai, E., Celeste, C. & Shafer, S. (2004). Digital Inequality: From Unequal Access to Differentiated Use: In K. Neckerman (Hrsg.), *Social Inequality.* Sage, S. 355–400.

Elsherbiny, A. (2020). *Driven by the Age of Automation: How Multinational Companies Fuel Diversity & Inclusion through Technology in the 21st Century.* Unveröffentlichte Masterarbeit, RWTH Aachen University.

Flake, R., Kochskämper, S., Risius, P. & Seyda, S. (2018). Fachkräfteengpass in der Altenpflege. IW-Trends 3/2018. https://www.iwkoeln.de/fileadmin/user_upload/Studien/IW-Trends/PDF/2018/IW-Trends_2018-03-02_Pflegefallzahlen.pdf [Zugriff: 8.6.2021]

Flecker, J. (2016). *Space, Place and Global Digital Work.* Palgrave MacMillan.

Frey, R. (2020). *Geschlecht und Gewalt im digitalen Raum. Eine qualitative Analyse der Erscheinungsformen, Betroffenheiten und Handlungsmöglichkeiten unter Berücksichtigung intersektionaler Aspekte. Expertise für den Dritten Gleichstellungsbericht der Bundesregierung.* https://www.damigra.de/wp-content/uploads/Expertise-Geschlecht-und-Gewalt-im-digitalen-Raum-20201.pdf [Zugriff: 18.5.2021].

Hasenkamp, M. (2020): *#MeToo and the Power of Social Media. Hashtag Aktivismus zwischen Solidarität und Polarisierung? Vortrag im Rahmen der Ringvorlesung „Gender und Diversity entdecken. Kulturwandel an Hochschulen*“, online am 16. November 2020, Gender Campus der Otto von Guericke Universität Magdeburg.

Hummel, D., Stieß, I. & Sauer, A. (2020). Technikfolgenabschätzung und Geschlecht: Bestandsaufnahme und Identifizierung von Diskursschnittstellen mit besonderem Fokus auf Digitalisierung. Expertise für den Dritten Gleichstellungsbericht der Bundesregierung, https://www.dritter-gleichstellungsbericht.de/de/article/228.technikfolgenabsch%C3%A4tzung-und-geschlecht-bestandsaufnahme-und-identifizierung-von-diskursschnittstellen-mit-besonderem-fokus-auf-digitalisierung.html [Zugriff: 18.5.2021].

Initiative D21 e. V. & Kompetenzzentrum Technik-Diversity-Chancengleichheit e. V. (Hrsg.) (2020). Digital Gender Gap. Lagebild zu Gender(un)gleichheiten in der digitalisierten Welt. https://initiatived21.de/app/uploads/2020/01/d21_digitalgendergap.pdf [Zugriff: 29.4.2021].

Kesselring, S. & Vogl, G. (2010). *Betriebliche Mobilitätsregime. Die sozialen Kosten mobiler Arbeit.* edition sigma.

Kompetenzzentrum Technik-Diversity-Chancengleichheit e. V. (2021): Datentool. Informatik. https://www.kompetenzz.de/service/datentool [Zugriff: 29.4.2021].

Krüger, H. (2009). Gesellschaftsanalyse: der Institutionenansatz in der Geschlechterforschung. In: G.-A. Knapp & A. Wetterer (Hrsg.), Sozi*ale Verortung der Geschlechter. Gesellschaftstheorie und feministische Kritik.* 4. Aufl. Westfälisches Dampfboot, S. 63–90.

Kutzner, E. & Schnier, V. (2017). Geschlechterverhältnisse in Digitalisierungsprozessen von Arbeit. *Arbeit,* 26(1), 137–157.

Lott, Y. (2015). Costs and benefits of flexibility and autonomy in working time: The same for women and men? WSI-Diskussionspapier 196 (2/2015). Hans-Böckler-

Stiftung, Wirtschafts- und Sozialwissenschaftliches Institut. www.econstor.eu/bitstream/10419/107636/1/819011991.pdf [Zugriff: 1.4.2021].

Marrs, K., Bultemeier, A. & Boes, A. (2015). Wo geht die Reise hin? Digitaler Umbruch von Arbeit – neue Möglichkeitsräume für Frauen? In *Frau geht vor, DGB-Infobrief 3/2015*, Berlin, Frankfurt, S. 4–6. http://frauen.dgb.de/themen/++co++ab52a410-8e05-11e5-a333-52540023ef1a [Zugriff: 1.4.2021].

Mau, S. (2017). *Das metrische Wir. Über die Quantifizierung des Sozialen.* Suhrkamp.

Meier-Gräwe, U. (2016). *Wenn Frauenbranchen auf Plattformen auswandern:* Helping & Co., Vortrag im Rahmen der 3. Gender Studies Tagung „Arbeit 4.0 – Blind Spot Gender" des Deutschen Instituts für Wirtschaftsforschung in Kooperation mit der Friedrich-Ebert-Stiftung am 22.9.2016 in Berlin. www.diw.de/documents/dokumentenarchiv/17/diw_01.c.543562.de/meier-graewe_genderstudiestagung2016.pdf [Zugriff: 1.4.2021].

Oliveira, D. (2017). *Gender und Digitalisierung. Wie Technik allein die Geschlechterfrage nicht lösen wird.* Reihe Hans-Böckler-Stiftung, Working Paper Nr. 37. Düsseldorf: HBS.

Panova, R., Sulak, H., Bujard, M. & Wolf, L. (2017). Die Rushhour des Lebens im Familienzyklus: Zeitverwendung von Männern und Frauen. In: Statistisches Bundesamt (Hrsg.), *Analysen zur Zeitverwendung in Deutschland. Beiträge zur Ergebniskonferenz der Zeitverwendungserhebung 2012/2013* am 5./6. Oktober 2016 in Wiesbaden, S. 45– 65. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Einkommen-Konsum-Lebensbedingungen/Zeitverwendung/Publikationen/Downloads-Zeitverwendung/tagungsband-wie-die-zeit-vergeht-5639103169004.pdf?__blob=publicationFile [Zugriff: 16.03.2021].

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information.* Harvard University Press.

Prietl, B. (2019). Die Versprechen von Big Data im Spiegel feministischer Rationalitätskritik. In *Gender* 3/2019, 11–25.

Rommes, E., Bath, C. & Maass, S. (2012). Methods for Intervention. Gender Analysis and Feminist Design of ICT. In *Science, Technology & Human Values* 37(6), 653–662.

Sachverständigenkommission für den Dritten Gleichstellungsbericht der Bundesregierung (2021). Digitalisierung geschlechtergerecht gestalten. Gutachten für den Dritten Gleichstellungsbericht der Bundesregierung. Berlin: Geschäftsstelle Dritter Gleichstellungsbericht. www.dritter-gleichstellungsbericht.de/gutachten3gleichstellungsbericht.pdf [Zugriff: 01.04.2021].

Schildmann, C. (2018). *Geschlechtergerechtigkeit in der digitalen Arbeitswelt. Fünf entscheidende Felder.* böll.brief TEILHABEGESELLSCHAFT #6. Berlin: HBS.

Stettes, O. (2016). *Arbeitswelt der Zukunft. Wie die Digitalisierung den Arbeitsmarkt verändert.* IW-Analysen Nr. 108. Köln: IDW.

The Economist (2017). Silicon Valley's sexism problem. Venture capitalists are bright, clannish and almost exclusively male. www.economist.com/news/leaders/21720621-venture-capitalists-are-bright-clannish-and-almost-exclusively-male-silicon-valleys-sexism [Zugriff: 18.05.2021].

Van Oost, E. (2003). „Materialized Gender". How Shavers Configure the Users' Femininity and Masculinity. In N. Oudshoorn & T. Pinch (Hrsg.), *How Users Matter. The Co-Connstruction of Users and Technology.* MIT Press, S. 193–208.

Weber, L. (2020). Digitalisierung, Geschlechtliche Zuweisungsprozesse und De/Professionalisierung in der Care-Arbeit. In K. Becker, K. Binner & F. Décieux (Hrsg.),

Gespannte Arbeits- und Geschlechterverhältnisse im Marktkapitalismus. Geschlecht und Gesellschaft. Springer Fachmedien Wiesbaden, S. 55–78.

Will-Zocholl, M. & Clasen, E. (2020). IngenieurInnen@work: Geschlechterperspektiven auf die Digitalisierung von Wissensarbeit im Engineering. In: K. Becker, K. Binner & F. Décieux (Hrsg.), *Gespannte Arbeits- und Geschlechterverhältnisse im Marktkapitalismus.* Geschlecht und Gesellschaft. Springer Fachmedien Wiesbaden, S. 99–122.

Opake Systeme künstlicher Intelligenz und das Problem der Verantwortungslücke

Christoph Sebastian Widdau / Carlos Zednik

Einleitung

Menschen können etwas herstellen, das als „autonom“ und „selbstlernend“ bezeichnet wird. Sie sind in der Lage, Artefakte zu konstruieren, die nicht nur Sachen und Instrumente sind, sondern die selbsttätig vorgehen, als lernfähig gelten und so komplex sind, dass man sie in ihren Wirkweisen nicht vollständig durchschauen kann. Kurzum: Sie sind fähig, opake Systeme künstlicher Intelligenz zu produzieren. Beispielsweise ist ein vollautomatisiertes Fahrzeug, das von künstlichen neuronalen Netzen betrieben wird, ein solches System. Es soll den Nutzenden ermöglichen, sich nach einer Zieleingabe ohne weitere eigene Eingriffe von einem zum anderen Ort fortbewegen zu lassen. Als Passagierin oder Passagier des vollautomatisierten Fahrzeugs wird man ohne menschliche Fahrerin oder menschlichen Fahrer transportiert. Damit es funktionieren kann, ist der Gang von Entwicklungsschritten, Programmierungen, automatisierten Lernprozessen, Tests und Prüfverfahren nötig. So soll das Fahrzeug lernen, die Regeln des Straßenverkehrs anzuwenden und die Straßenlinienführung erkennen zu können; ansonsten wäre es nicht zweckdienlich. Es soll in verschiedenen Situationen und Konstellationen agieren und reagieren, auch, um mögliche Unfälle nicht entstehen zu lassen. Weil ein solches Fahrzeug vom Menschen allerdings nicht so vorprogrammiert und eingestellt werden kann, dass es alle möglichen Situationen und Konstellationen zu erfassen und zu verarbeiten vermag, werden Methoden des maschinellen Lernens eingesetzt, die es dem Fahrzeug ermöglichen, sich selbst zu programmieren. Diese Methoden basieren auf automatisch erzeugten, komplexen Wechselwirkungen zwischen dem Fahrzeug und seiner Umgebung, die es dem Fahrzeug erlauben, sich Verhaltensweisen wie das sichere und gesetzeskonforme Fahren

anzueignen. Weil diese Methoden jedoch so komplex sind und autonom angewendet werden, haben sie zur Folge, dass Programmierende nicht immer in der Lage sind, vorauszusehen, wie sich das Fahrzeug in jeder Situation verhalten wird, und in der Regel nicht nachvollziehen können, wieso es sich gerade so verhält und nicht anders.

Opake Systeme künstlicher Intelligenz wie ein von künstlichen neuronalen Netzen betriebenes vollautomatisiertes Fahrzeug sind keine einfachen Werkzeuge und keine einfachen Automaten. Sie sind komplex, reaktiv, womöglich einst sogar selbst absichtsvoll aktiv. Ein Hammer ist dies nicht: Er ist ein einfaches Werkzeug. Er wurde handwerklicher Zwecke wegen konstruiert und soll zur Zweckerfüllung taugen. Selbstverständlich kann man ihn zweckentfremden und zum Element eines musikalischen Arrangements machen. Oder man kann ihn nutzen, um jemanden zu attackieren. Wie auch immer man ihn nutzt: Seine Nutzung und deren Folgen sind unter Normalbedingungen ganz und gar abhängig von demjenigen, der den Hammer hält und bewegt. Der Hammer ist nicht autonom und nicht selbstlernend. Weder tut er etwas, denn dazu ist er nicht in der Lage, noch kann er Eindrücke verarbeiten. Kommt ein Mensch durch einen Hammerschlag zu Tode, wird niemand, der vernünftig ist, den Gedanken haben, den Hammer in die Verantwortung zu nehmen und ihn Rede und Antwort stehen lassen zu wollen; auch ist dies so bei einem einfachen Automaten, wenngleich er, ist er erst einmal in Gang gesetzt, mechanisch Tätigkeiten verrichtet. Wird ein Mensch durch einen solchen Automaten, beispielsweise einen Kaffeeautomaten, verletzt, dann liegt es daran, dass dieser schlecht eingerichtet oder programmiert, nicht sorgfältig überholt, fehlerhaft eingesetzt oder zweckentfremdend mit ihm umgegangen worden ist. Hinsichtlich des Gebrauchs des Hammers und des Kaffeeautomaten stellen sich, sobald sich Fragen nach der *Verantwortung* stellen, diese nicht ihnen, sondern denen, die ihre Entwicklung und Produktion erlaubt und forciert haben, und denen, die sie nutzen und warten.

Aufgrund der Entwicklung und des Einsatzes von opaken Systemen künstlicher Intelligenz ergeben sich moralische Probleme, die sowohl in den Akademien als auch in der breiten Öffentlichkeit Aufmerksamkeit erregen. Ein bereits genanntes ist das Problem der Verantwortung für Aktionen dieser Systeme und deren Folgen. Um dies einleitend und exemplarisch zu veranschaulichen: Wenn ein vollautomatisiertes Fahrzeug sich selbst so lenkt und beschleunigt, dass es entgegen der Absicht, Eingabe und Erwartung seiner Entwickelnden, Vermarktenden und Nutzenden Passantinnen und Passanten überfährt und tötet, wer oder was ist dann für deren „Zu-Tode-Kommen“ moralisch verantwortlich? Die Entwickelnden haben das sogenannte autonome und selbstlernende System nicht so programmiert, dass es jemanden überfahren soll. Ereignis, Aktion und Folge waren von ihnen, die Absichten haben können, nicht beabsichtigt. Sie haben das von ihnen geschaffene System allerdings auch nicht vollständig „in der Hand“, eben weil es opak, autonom, und selbstlernend

ist. Das Fahrzeug selbst wird sich jedoch keiner Richterin und keinem Richter stellen müssen.[1] Wenigstens Angehörige der getöteten Passantinnen und Passanten werden aber nicht nur die Frage nach der kausalen, sondern auch die nach der moralischen sowie, darüber hinaus, der rechtlichen Verantwortung stellen: „Wer oder was ist für den Tod des von uns geliebten Menschen verantwortlich? Wer hat sich Schuld aufgeladen? Wer oder was ist dafür haftbar zu machen? Wer oder was hat darüber Rechenschaft abzulegen?“ Sind es doch die Entwickelnden? Die Vermarktenden? Die Nutzenden? Ist es die Institution des Gesetzgebers, die die Entwicklung und Nutzung zugelassen hat? Sind es die Geldgebenden, ohne deren Mittel die Entwicklung nicht möglich gewesen wäre? Jemand anders? Niemand?

Es gibt Autorinnen und Autoren, die behaupten, dass die Antwort „Niemand“ die richtige ist: In der Ethik der künstlichen Intelligenz wird hinsichtlich der Verantwortungsfrage manchmal die These der *Verantwortungslücke* vertreten, die seit fast zwanzig Jahren umstritten ist (Matthias 2004, Champagne & Tonkens 2015, Tigard 2021).[2] Wenn jemand davon spricht, dass es in einer Situation eine Verantwortungslücke gibt, dann meint er oder sie dies: Eine Handlung (bzw. eine Tätigkeit) oder eine Handlungskonsequenz, für die jemand oder etwas moralisch verantwortlich gemacht werden sollte, ist derart beschaffen, dass niemand oder nichts moralisch verantwortlich gemacht werden *kann.* Sie besagt, exemplarisch gewendet, dass in dem Fall des vollautomatisierten Fahrzeugs und der getöteten Passantinnen und Passanten nichts und niemand moralisch verantwortlich ist oder gemacht werden kann. Warum? Weil in diesem Fall nichts und niemand die Kriterien erfüllt, moralisch verantwortlich zu sein. Freilich gab es politische und rechtliche Bedingungen, die die Entwicklung des Fahrzeugs ermöglicht haben. Freilich bedurfte es einer oder eines Geldgebenden, um seine Entwicklung zu forcieren. Freilich hat es jemand entwickelt. Freilich hat es jemand eingesetzt und genutzt. Freilich hat es unbeabsichtigt und unerwünscht reagiert. Dennoch ist das Fahrzeug selbsttätig und in nicht zureichend prognostizierbarer Weise agierend. Es ist eben kein einfaches Werkzeug und kein einfacher Automat: Es ist ein opakes, selbsttä-

1 Akademische Debatten über die Möglichkeit, dass Roboter als moralische und rechtliche Subjekte gelten könnten, werden nichtsdestoweniger geführt (vgl. Ashrafian 2014). In dem vorliegenden Beitrag gehen wir davon aus, dass zumindest bis heute keiner der bislang entwickelten Roboter einen solchen Status hat.

2 Santoni de Sio und Mecacci (2021) gehen sogar davon aus, dass es mehrere Verantwortungslücken (*Responsibility Gaps*) gibt. Ein Beispiel sei die *Active Responsibility Gap.* Die Autoren schreiben: „This gap consists in the risk that persons designing, using, and interacting with AI [= *Artificial Intelligence*] may not be sufficiently aware, capable, and motivated to see and act according to their moral obligations towards the behaviour of the systems they design, control, or use. In particular, this gap concerns the obligation to ensure that these systems do not impact negatively on the rights and interests of other persons and, ideally, positively contribute to their well-being instead.“ Wir begreifen dies (es ist ein Beispiel neben anderen) nicht als eine Verantwortungslücke, sondern als ein Problem unzureichend wahrgenommener Verantwortung.

tiges System. Aufgrund seiner Opazität und Autonomie, so sagen es die Vertretenden der These der Verantwortungslücke, sind die Institution des Gesetzgebers, Geldgebende, Entwickelnde und Nutzende nicht moralisch verantwortlich. Denn auch wenn diese Akteurinnen und Akteure das Verhalten des Fahrzeugs vorhersagen oder nachvollziehen wollten, unterminiert seine Opazität und Autonomie die Möglichkeit, das zu tun. Und das opake System künstlicher Intelligenz ist wiederum selbst kein Verantwortungsträger in einem moralisch gehaltvollen Sinn (es kann sich und sein Verhalten nicht verantworten). Die Tötung der Passantinnen und Passanten gleicht eher, so könnten Vertretende der These versucht sein, sie zu rechtfertigen, einem tragischen Ereignis oder einem schicksalhaften Geschehen als einem Fall moralisch relevanten Handelns, das es zu bewerten und zu sanktionieren gilt. Und für ein Ereignis wie ein nichtanthropogenes Erdbeben und seine Folgen wird man niemanden moralisch verantwortlich machen.[3] Wir vertreten in diesem Beitrag die Gegenthese: *Es gibt in solchen Fällen wie dem des vollautomatisierten Fahrzeugs und der getöteten Passantinnen und Passanten keine Verantwortungslücke, deretwegen niemand die moralische Verantwortung tragen kann und sollte.* Stattdessen gibt es geteilte Verantwortung im kollektivethischen Sinne – die Natur des opaken Systems künstlicher Intelligenz bedingt keine moralisch relevante Lücke, sondern erzwingt normativ die Verantwortungszuschreibung und Haftbarmachung derer, die ein solches System in die Welt bringen, bringen lassen und einsetzen. Auch wenn es, wie stets in Fragen kollektiver Verantwortung[4], sehr schwierig ist, genau zu sagen, wer für etwas Bestimmtes moralische Verantwortung hat, darf dies nicht zu der falschen Annahme führen, dass sie niemand hat.[5] Man kann begründen, dass sie jemand hat und dass wir dann prüfen müssen, wer sie hat und in welchem Maße. Wir beschränken uns in dem vorliegenden Beitrag darauf, zu zeigen, dass die

3 Gewiss: Gottesgläubige könnten versucht sein, Gott (oder Götter) für das nichtanthropogene Erdbeben verantwortlich zu machen. Wir werden dieser (allzu spekulativen) Annahme hier nicht nachgehen.

4 Ein Fall kollektiver Verantwortung ist der des anthropogenen Klimawandels, bei dem das ethische Problem der vielen Hände (*Problem of Many Hands*) besteht. Dieser Wandel ist menschengemacht. Aber es ist sehr schwierig, für ihn und die von ihm hervorgerufenen Folgen (zum Beispiel den globalen Anstieg des Meeresspiegels und die Überflutung bewohnter Inseln) *ganz genau* die kausal, moralisch und rechtlich Verantwortlichen zu bestimmen. Siehe zum Problem beispielsweise van de Poel et al. (2018).

5 Wenn es in einem Beitrag heißt „In bestimmten Situationen können autonome Systeme die Verantwortungszuschreibung unterminieren. Es wird unklar, wer Verantwortung trägt. Es entsteht eine ‚Verantwortungslücke', in der weder Programmierern noch Anwendern die volle Verantwortung zugeschrieben werden kann" (Misselhorn 2018, S. 96), dann verstehen wir dies, dass unklar ist, *wer genau* und wer *wie viel* Verantwortung hat, nicht als Beleg für eine Verantwortungslücke. Eine solche würde erst dann gerissen, wenn *niemand* verantwortlich gemacht werden *kann.* Die Unterscheidung zwischen voller und halber Verantwortung ist wichtig und analytisch relevant. Aber die These der Verantwortungslücke verstehen wir hier schärfer.

Verantwortungslücke nicht besteht. Mögliche Konsequenzen, die aus dieser Einsicht folgen, werden wir nicht diskutieren.

Kausale und moralische Verantwortung

In der deutschen Sprache unterscheiden wir unter anderem zwischen nicht-moralischer (im Sinne von bloß kausaler) und moralischer Verantwortung.[6] „Nicht moralisch verantwortlich zu sein" bedeutet hierbei, nur kausal für etwas verantwortlich zu sein. Wir sagen, wenn ein Drucker nicht tut, was er tun soll, etwas wie: „Verantwortlich dafür ist ein Fehler im System." Damit meinen wir: Wenn es den Fehler im System nicht geben würde, dann würde der Drucker tun, was er eigentlich tun soll. Wir erwarten aber von dem Fehler im System oder dem Drucker selbst nicht, dass er Rechenschaft über das Problem ablegen kann und dass er schuldhaft einer Pflicht, etwas Bestimmtes tun zu sollen, nicht nachgekommen ist. Bei der moralischen Verantwortung geht es genau darum: Ein Mensch, der sein Verhalten willentlich und absichtsvoll bestimmen kann, soll etwas tun und tut es nicht oder nur unzureichend. Deswegen darf erwartet werden, dass er Rechenschaft abzulegen hat und erklären soll, warum er nicht getan hat, was er hätte tun sollen. Oder, wenn man sich nicht sicher ist, was jemand hätte tun sollen, zu erklären, warum er dachte, dieses und nicht jenes tun zu sollen und entsprechend gehandelt hat. Beispielsweise sagen wir, wenn jemand nicht tut, was er tun soll, etwas wie: „Lewis hätte Alice nicht anlügen dürfen – es ist daher richtig, dass sie ihn zur Rede stellt." Oder wir sagen in Fällen, in denen wir nicht sicher sind, was er hätte tun sollen: „Es scheint mir problematisch zu sein, dass Lewis Alice angelogen hat – er sollte uns erklären, warum er das getan hat." Sicher ist: Lewis hat Alice angelogen. Kausal verantwortlich dafür, dass die Lüge in der Welt ist, ist Lewis. Eine Lüge auszusprechen, ist aber stets auch von moralischer Relevanz. Wie steht es dann um seine moralische Verantwortung?

Betrachten wir eine neue Situation, in der die moralische Verantwortung der Akteurin oder des Akteurs, die oder der kausal verantwortlich ist, eine Rolle spielen kann und spielt – hierbei fallen kausale *und* moralische Verantwortung bei einem Handelnden zusammen:

[S1] Wolfgang möchte sich einen Wagen kaufen, hat aber kein Geld. Sein Freund Siegfried hat es. Wolfgang bittet Siegfried darum, ihm Geld zu leihen, um sich einen Maserati kaufen zu können und verspricht ihm, die geliehene Summe zeitnah zu begleichen, obwohl er sich sicher ist, dass er dies nicht wird

6 Die philosophische Fachdiskussion über die Bestimmung des Begriffs der Verantwortung, seine Spielarten sowie die Relata und Strukturen der Verantwortung können wir an dieser Stelle nicht rekapitulieren. Um einen Überblick über die Fachdiskussion gewinnen zu können, sei empfohlen: Heidbrink et al. (2017).

tun können. Siegfried, der Wolfgangs finanzielle Situation nicht im Detail kennt, leiht ihm das Geld, ohne formelle Abmachung. Wolfgang kauft das Auto. Nach einem halben Jahr kommt Siegfried auf Wolfgang zu und fragt nach dem Geld. Wolfgang vertröstet, dass er es sicher bald haben werde. Siegfried wartet. Nach weiteren Monaten stellt er Wolfgang erneut zur Rede. Wieder blitzt er ab. Siegfried kündigt Wolfgang daraufhin die Freundschaft auf: „Du hattest nie die Absicht, mir die Summe zurückzugeben! Es ging dir nur um dein Interesse! Du hast mich hinters Licht geführt!“

Betrachten wir nun drei Situationen, in denen die moralische Verantwortung der Agierenden oder Ereignisse, die kausal verantwortlich sind, keine Rolle spielen kann (weil es sie vonseiten dieser Agierenden sowie Ereignisse nicht geben kann) – dies sind [S2], [S4] und [S5] –, sowie eine Situation, in der wir von einer begrenzten (oder halben oder geteilten) moralischen Verantwortung der kausal verantwortlichen Akteurin bzw. des kausal verantwortlichen Akteurs ausgehen dürfen – dies ist [S3].

[S2] Junior, der zweijährige Sohn von Arnold und Danny, wird während einer Party, die Freunde seiner Eltern ausrichten, von diesen alleingelassen. Die Eltern sagen sich: „Uns gehört die Nacht; der Kleine kommt schon allein zurecht – und was soll er schon anstellen?“ Junior krabbelt unbehelligt durch den Garten, jagt einem Glühwürmchen nach und rempelt dabei eine ältere Dame um, die rücklings zu ihm stand und ihn nicht sehen konnte. Die Dame schreit und stürzt. Arnold und Danny eilen, nun von weiteren Gästen gerufen, herbei. Die Dame fragt mit schmerzverzerrtem Gesicht, wie man denn ein so kleines Kind allein durch den Garten krabbeln und aus den Augen lassen konnte. Arnold und Danny schauen erst sich an und dann betreten zu Boden.

[S3] Der sechsjährige Lausbub Michel möchte einer Freundin eine Freude machen und sie überraschen. Er möchte ihr eine Botschaft ins Haus bringen. Er schreibt ihr einen kleinen Zettel, bindet diesen an einen Stein und wirft ihn, auf der Straße stehend, gegen und durch das Fenster, hinter dem er seine Freundin vermutet. Das Glas des Fensters bricht. Zum Glück wird niemand von dem Stein getroffen. Die Eltern der Freundin eilen aus dem Haus und fragen Michel, was er sich dabei gedacht hat. Er entschuldigt sich und sagt, dass er das, den Glasbruch, nicht gewollt habe; und seine Freundin wollte er natürlich auch nicht treffen. Michels Eltern wird eine wütende Rede der anderen Eltern erwarten (in Anbetracht der Sorge um ihre Tochter, die hätte getroffen werden können) und eine Rechnung über die Installationskosten eines neuen Fensters.

[S4] Rexi, die Schäferhündin von Leonora, entfernt sich während eines frühmorgendlichen Spaziergangs von ihrer Halterin. Normalerweise ist zu dieser Zeit niemand auf der Straße, sodass Leonora Rexi unangeleint lässt. Rexi ist auch eine wohlerzogene Hündin, die in der Regel menschenfreundlich ist. Rexi, nicht mehr in Leonoras Sichtfeld, trifft unerwartet auf den Spaziergänger Immanuel, der am Wegesrand steht und nichts tut, was Rexi normalerweise reizen könnte. Rexi springt aber unerwarteterweise an ihm hoch und fletscht

kurz mit den Zähnen, was Immanuel, der Angst vor Hunden hat, fast in Panik versetzt; gerade noch so kann er stillhalten und Rexi lässt wieder von ihm ab.

[S5] Kim liegt nachts in seinem Bett, als ihn ein schweres Erdbeben aus dem Schlaf reißt. Hastig eilt er aus dem Haus, um seine Haut zu retten, und muss mit ansehen, wie sein Haus in sich zusammenfällt. Das Erdbeben zerstört sein Obdach. Es traf ihn (und alle anderen von ihm Betroffenen in der näheren Umgebung) ohne Vorwarnung. Kim und die anderen Betroffenen hatten zuvor auch nichts getan, was das Erdbeben hätte verursachen können.

Warum spielt in [S1] die moralische Verantwortung des kausal verantwortlichen Akteurs eine Rolle, in [S2], [S4] und [S5] nicht, in [S3] bedingt? Wolfgang ist ein Wesen, das fähig ist, Entscheidungen zu treffen, zu handeln und mögliche Konsequenzen seines Handelns abzusehen. Er kann in dem geschilderten Fall auch ganz gut abschätzen, welche Konsequenzen wahrscheinlich eintreten werden. Er weiß auch, was es bedeutet, einander zu vertrauen. Er hätte den Wagen nicht kaufen müssen; er hätte Siegfried nicht um Geld bitten müssen. Wolfgang wurde zudem nicht von außen gezwungen, Siegfried um Geld zu bitten; er hätte dies auch nicht tun können, hatte hierfür einen Spielraum. Aber freiwillig und absichtsvoll bat er um Geld. Wolfgang wurde nicht willenlos gemacht, nicht unter schwerste Drogen gesetzt oder einer Gehirnwäsche unterzogen – er war voll zurechnungsfähig. Und Wolfgang hätte, nachdem ihm Siegfried das Geld gegeben hat, immer noch zurückziehen können. Er hätte Siegfried den Fehler, ihn so um Geld zu bitten, eingestehen können. Er hätte auch den Wagen (mit Verlust) zurückgeben und Siegfried zumindest einen Teil der Geldsumme erstatten können. Wolfgang hätte jederzeit etwas tun können, das vielleicht zur Rettung der Freundschaft beigetragen hätte. Wolfgang ist in der Lage, seine moralisch relevante Entscheidung und seine Handlung zu begründen – ob er eine normativ gute Begründung findet, ist eine andere Frage. Wolfgang ist prinzipiell in der Lage, Rede und Antwort zu stehen und sich zu rechtfertigen und zu kritisieren. Angesichts dessen, dass er Siegfried zwei Übel zugefügt hat – zum einen hat er ihn um das Geld gebracht, zum anderen hat er seinen Glauben an die Freundschaft, ein Vertrauensverhältnis, erschüttert –, liegt der Bedarf an einer Rechtfertigung ethisch auf der Hand. Und dass er selbst derjenige ist, der sich in diesem Fall zu rechtfertigen hat, steht außer Frage.

Vieles von dem, was Wolfgang in [S1] kann oder gekonnt hätte, können und könnten Junior in [S2], Rexi in [S4] und das Erdbeben in [S5] nicht. Offensichtlich sind Junior, Rexi und das Erdbeben jeweils kausal verantwortlich für den Sturz der Dame, die Angst des Spaziergängers und das Zerstören des Hauses. Sie sind in erster Linie diejenigen, ohne die eine bestimmte Folge nicht eingetreten wäre. Aber man würde sie nicht (und zwar gar nicht) als moralisch verantwortlich ansehen. Junior entscheidet sich nicht dafür, die ältere Dame stürzen zu lassen. Er versteht nicht, was das heißt. Er würde es auch nicht verstehen können, wäre jemand bereit, ihm dafür einen Vorwurf zu machen. Er

weiß nicht, was ein Vorwurf ist und warum man jemandem Vorwürfe machen kann. Weder kann Junior abwägen und urteilend entscheiden, was er tut, noch weiß er, was er tut, noch kann er vernünftig absehen, was eine bestimmte Tat zur Folge haben wird. Ihm mangelt es an den Eigenschaften und Fähigkeiten, moralisch verantwortlich sein zu können. Junior wird, wenn er sich gewöhnlich entwickelt, einst ein Verantwortungsträger sein. Zurzeit jedoch ist er dies nicht. Die Hündin Rexi und das Erdbeben, ein nichtmenschliches Tier und ein Naturereignis, haben solche Eigenschaften und Fähigkeiten nicht und werden sie niemals haben. Ihnen mangelt es im Gegensatz zu Junior an jedem Potenzial, sie haben zu können.

In den Szenarien [S1], [S2] und [S4] entsteht keine Verantwortungslücke; in [S5] hingegen schon. Siegfried wird sich in [S1] an Wolfgang wenden, um entschädigt zu werden und wenn dies nicht hilft, sich an Dritte wenden, um zu prüfen, ob man Wolfgang vonseiten des Rechts zu einer Entschädigung zwingen kann: Der kausal Verantwortliche ist auch, wie oben bereits geschildert, der moralisch Verantwortliche und der, den man gegebenenfalls auch rechtlich verantwortlich zu machen sucht. Es gibt in [S1] keine Verantwortungslücke.

Die ältere Dame in [S2] wird sich an die Eltern von Junior, an Arnold und Danny, wenden, um sich zu beschweren und gegebenenfalls Geld für die Arztkosten zu erhalten, im Notfall rechtlich zu erstreiten: Nicht der kausal Verantwortliche ist auch der moralisch Verantwortliche, sondern die Aufsichtspersonen des in diesem Fall kausal Verantwortlichen werden als moralisch (und rechtlich) Verantwortliche angesprochen, und es ist zu prüfen, ob sie den Schaden hätten verhindern können und verhindern sollen. Schadensersatz sollen *sie* leisten, nicht Junior: Die Eltern werden angesichts ihrer behaupteten Verantwortungslosigkeit gescholten und für Juniors Verhalten haftbar gemacht. Auch in [S2] gibt es keine Verantwortungslücke.

Immanuel in [S4] wird die Halterin Leonora, wenn sie sich Rexi und ihm nähern wird, einen Vorwurf machen, nämlich den, dass die Hündin unangeleint ist und fremde Menschen anspringt: Nicht die für das Erschrecken unmittelbar kausal verantwortliche Hündin ist moralisch verantwortlich, sondern die Hundehalterin wird in die Pflicht genommen. Hätte sie, als moralisch Verantwortliche, in erster Linie anders gehandelt (Anleinen), hätte die Hündin in zweiter Linie nicht Immanuel anspringen können.[7] Auch in [S4] gibt es keine Verantwortungslücke.

7 Gibt es, wie in manchen Ländern, eine allgemeine Leinenpflicht für Hunde in der Öffentlichkeit, haben wir es hierbei mit einem Fall zu tun, in dem (eines ethischen Problems wegen, nämlich des Problems der Durchsetzung des als angemessen erachteten Zusammenlebens von Hundehaltern, Nichthundehaltern und den nichtmenschlichen Tieren) eine rechtliche Sanktion (ein Bußgeld oder Anderes) drohen könnte. Leonora könnte dann rechtlich zur Verantwortung gezogen werden.

In [S5] hingegen schon. Kim wird *niemandem* und *nichts* einen Vorwurf machen können, niemanden zur Rechenschaft ziehen, das kausal Verantwortliche (das Erdbeben) nicht als Objekt der Vergeltung erachten können, auch nichts an dessen Stelle setzen können, sondern vielleicht an seinem Schicksal leiden und sich an seine Versicherung wenden, um zu prüfen, ob ihm aufgrund seiner vorgängigen Vertragsschlüsse eine Entschädigung zusteht (angesichts eines Schadens, für den niemand und nichts moralisch verantwortlich ist).

Wie verhält es sich in [S3]? Einiges von dem, was Wolfgang in [S1] kann oder gekonnt hätte, kann Michel in [S3] nicht, manches noch unvollkommen. Michel ist sich seiner selbst bewusst. Er kann sprechen. Er hat ein Verständnis davon, was es heißt, miteinander umzugehen. Er wurde bis zum sechsten Lebensjahr nicht nur am Leben erhalten, sondern auch mindestens in Ansätzen erzogen. Michel kann weniger gut als Wolfgang abschätzen, welche Folgen seine Handlungen haben werden, aber er kann es. Michel weiß auch nicht so gut, was es bedeutet, jemandem einen Sachschaden oder einen anderen Schaden zuzufügen. Ihm mangelt es (noch) an Erfahrungen und an der Ausprägung eigener Urteilskraft. Allerdings kann Michel bereits deutlich mehr als Junior in [S2]: Er ist schon *weiter auf dem Weg*, ein Akteur zu werden, dem man uneingeschränkt moralische Verantwortung wird zusprechen können. Der Sechsjährige darf im Dorf, das er kennt, allein spazieren und seine Freundin besuchen. Er ist dazu in der Lage, und es ist ihm erlaubt. Zugleich hätten seine Eltern ihm womöglich bereits zuvor mehr über Schäden an Leib und Leben sowie Sachschäden und deren Zustandekommen beibringen können, womöglich sollen. Man kann sagen, dass die kausale Verantwortung für die Tat bei Michel lag, die moralische Verantwortung jedoch bei Michel *und* seinen Eltern. Auch hier gibt es also keine Verantwortungslücke, auch wenn es sehr starke mildernde Umstände für beide Parteien, Michel und die Eltern, gibt, die in einer ethischen Beurteilung ins Gewicht fallen müssen.

Der Fall der opaken, autonomen künstlichen Intelligenz

Was ist in Situationen der Fall, in denen ein opakes System künstlicher Intelligenz agiert und für etwas kausal verantwortlich ist, für das man es eigentlich auch moralisch verantwortlich machen möchte? Nehmen wir, um dies zu veranschaulichen, an, dass sich dies ereignet:

[S6] Durch den Schuss der autonomen, selbstlernenden Waffe Killbot 3000, die in einem Krieg zweier verfeindeter Staaten eingesetzt wird und der zentrale Regeln des Kriegsrechts einprogrammiert sind, kommt ein vierjähriges Kind, also ein Mensch, der nicht zu den Kampftruppen gehört, also kein Kombattant ist, zu Tode. Dies ist ein Bruch mit den Regeln des Kriegsrechts. Es gab auch

keine unübersichtliche Lage: Das Kind war eigentlich als Kind zu identifizieren (und nicht etwa mit einem kleinwüchsigen Erwachsenen zu verwechseln) und es gab zurzeit der Tötung des Kindes keine Kampfhandlung, in der der Roboter involviert war. Es ist naheliegend, dass die Eltern des Kindes den, der für den Tod des Kindes verantwortlich ist, irgendwie zur Rechenschaft ziehen wollen.

Der Killbot 3000 hat in der Situation autonom agiert. Kein Mensch hat einen Knopf gedrückt, dessentwegen Killbot 3000 geschossen hat. Als opakes System war es für Außenstehende auch nicht vorhersehbar gewesen, dass Killbot 3000 so agieren wird. Und es ist für die Außenstehenden zurzeit des Schusses und in der Zeit nach dem Schuss kaum möglich, genau zu erklären, warum Killbot 3000 geschossen hat. Durch seine selbstlernende, autonome Funktionsweise kennt niemand die genauen Umstände, unter denen die Waffe feuert, noch können Außenstehende die Prozesse nachvollziehen, nach denen die Entscheidung zu feuern getroffen wird.

Wer oder was ist in [S6] moralisch verantwortlich? Ist hier überhaupt jemand oder etwas in diesem Sinne verantwortlich? Diejenigen, die die These der Verantwortungslücke vertreten, sagen, dass niemand und nichts moralisch verantwortlich ist. Eine Analyse von [S6] zeigt dies:

Erstens hätte das Kind nicht durch den Killbot 3000 zu Tode kommen können, hätte ihn nicht jemand entwickelt und die Entwicklung angeordnet und finanziert.

Zweitens wäre das Kind nicht durch den Killbot 3000 zu Tode gekommen, hätte man ihn nicht in dem Krieg eingesetzt.

Drittens ist das Kind durch den Killbot 3000 zu Tode gekommen, obwohl die Entwickelnden zumindest einiges daran gesetzt haben, dass derlei nicht passiert.

Viertens ist das Kind durch den Killbot 3000 zu Tode gekommen, weil dieser in einer Weise agiert hat, die so nicht vorauszusehen gewesen und (bislang) kaum nachzuvollziehen ist. Das Kind ist durch einen Akteur zu Tode gekommen, dessen Herstellende grundsätzlich nicht in der Lage sind, ihn vollumfänglich einzusehen, zu verstehen und sein Agieren vorauszusehen.

Fünftens ist das Kind durch ein opakes System zu Tode gekommen, das selbst nicht zur Verantwortung fähig ist, weil es nicht versteht, was es heißt, Verantwortung zu übernehmen, nicht moralische Werte bestimmen und moralisch evaluieren kann, die moralische Relevanz seiner Aktionen nicht nachempfinden kann und nicht Rechenschaft ablegen kann über das, was es getan hat: Man kann den Killbot 3000 überhaupt nicht moralisch zur Verantwortung ziehen.

Im Sinne der nichtmoralischen kausalen Verantwortung ist der Killbot 3000 in [S6] sehr wohl verantwortlich: Wenn er nicht geschossen hätte, dann wäre das Kind nicht zu Tode gekommen; weil er gefeuert hat, ist das Kind zu Tode gekommen. Dies ist unstrittig. Strittig ist, ob jemand im Sinne der moralischen Verantwortung verantwortlich ist. Es ergeben sich hierzu vier mögliche Thesen:

Erste These: Nichts und niemand ist für den Tod des Kindes moralisch verantwortlich.
Zweite These: Der Killbot 3000 ist für den Tod des Kindes moralisch verantwortlich.
Dritte These: Jemand, ein Mensch (oder eine Menge von Menschen), ist für den Tod des Kindes moralisch verantwortlich.
Vierte These: Der Killbot 3000 *und* jemand sind für den Tod des Kindes gemeinsam moralisch verantwortlich.

Die erste These ist die der Verantwortungslücke. Sie wäre widerlegt, wenn eine der drei anderen Thesen richtig ist. Die zweite These kann angesichts dessen, dass der Killbot 3000 nicht die Kriterien erfüllt, moralisch verantwortlich zu sein, nicht richtig sein. Er ist noch nicht einmal so weit, wie es Michel in [S3] ist. In bestimmtem Sinne gilt der Killbot 3000 als autonom, intelligent und selbstlernend. Aber zumindest noch gilt unter anderem dies: Er kann nicht moralisch evaluieren. Er kann sich nicht rechtfertigen. Er hat und kennt keine Schuld. Ihm kann man nichts vergelten. Er kennt kein eigenes und fremdes Übel. Er hat überhaupt keine Urteilskraft. Er kennt keine moralische Unsicherheit. Nehmen wir an, dass dies stimmt, dann kann die zweite These nicht richtig sein, und dann kann auch die vierte These nicht richtig sein. Die Kandidatin, die übrig bleibt, um die These der Verantwortungslücke zurückzuweisen, ist die dritte These. Ihr zufolge ist *jemand* für den Tod des Kindes moralisch verantwortlich – auch wenn es eventuell strittig ist, wer genau dies ist.

Wenn die erste These richtig wäre, dann hätten wir es mit einer Ähnlichkeit der Zurechenbarkeit von Verantwortung in den Fällen [S5] und [S6] zu tun. In [S5] wie in [S6] würde es dann eine Verantwortungslücke geben. Aber diese Ähnlichkeit ist grundsätzlich nicht gegeben. Denn das nichtanthropogene Erdbeben ist ein Naturereignis ohne Zweck, ohne Absicht, ohne Herstellende, ohne Zulassende. Der Killbot 3000 hingegen ist ein Artefakt mit programmierten Zwecken, absichtsvoll produziert, mit Herstellenden und einer zulassenden Autorität. Ein Artefakt wird von jemandem, der es bestimmter Zwecke wegen ersonnen hat und nutzen möchte, in die Welt gebracht. Besondere Artefakte wie der Killbot 3000 haben angesichts ihrer Lernfähigkeit und Autonomie die (vielleicht) unangenehme Eigenschaft, immer unabhängiger zu werden. Gewiss bleibt allerdings: Jemand hat ihn absichtlich hervorgebracht. Beim Erdbeben ist dies nicht der Fall. Alles Gemachte lässt sich auf Machende und deren Absichten zurückführen. Und für Mängel des Gemachten, beispielsweise, wird man in erster Linie die Machenden in die Verantwortung nehmen. Das Erdbeben ist ein schicksalhaftes Ereignis, das wir nicht in der Hand haben und haben können; der Killbot 3000 ist ein gemachtes Werk, das wir zumindest in Teilen in der Hand haben und gar nicht erst hätten in die Welt bringen müssen. Weil es opak, autonom und selbstlernend ist, haben wir es

nicht (mehr) vollständig in der Hand, aber es ist auch nicht etwas vollends Unbeherrschbares und Schicksalhaftes sowie nichts, was hinsichtlich seiner Entstehung nicht hätte verhindert werden können.

Der Killbot 3000 ist zweifellos ein besonderes Artefakt. Er ist kein Naturereignis (*Erdbeben*), kein einfaches Werkzeug (*Hammer*), kein autonomer moralischer Akteur (*Wolfgang*), kein Akteur, von dem wir mit guten Gründen annehmen dürfen, dass er einst ein moralischer Akteur sein wird (*Junior*, in weiterer Ferne, und *Michel*, in näherer Ferne) und kein Wesen, von dem wir wissen, dass es heute und einst ein nichtmoralischer Akteur sein wird (*die Hündin Rexi*).

Bei allen Ungereimtheiten kommt der Fall [S6] am ehesten den Fällen [S2] und [S4] nahe. Killbot, Kind und Hund tun etwas von sich aus. Sie haben kein Bewusstsein von dem, was sie tun und können es nicht moralisch bewerten. Sowohl das zweijährige Kind als auch der Hund sind Naturwesen. Hierin liegt offensichtlich ein Unterschied zwischen dem zweijährigen Kind und dem Hund auf der einen und dem Killbot 3000 auf der anderen Seite. Das zweijährige Kind ist ein durch eine sexuelle Handlung seiner Eltern in die Welt Gebrachtes, das, solange es selbst kein moralischer Akteur ist, unter der Obhut, Erziehung und Aufsicht der Eltern steht, weil sie es in die Welt gebracht haben und wissen konnten, dass es Vorhersehbares und Nichtvorhersehbares tun wird, ohne in seinen ersten Lebensjahren genau zu wissen, was es tun wird. Junior hat davon noch überhaupt keinen Begriff (im Gegensatz zu Michel). Er macht Vorhersehbares und Nichtvorhersehbares. Sowohl für das eine als auch für das andere können Eltern, wenn sie beispielsweise ihre Aufsichtspflicht verletzen, moralisch verantwortlich und haftbar gemacht werden. Auch ist dies so im Fall der Hundehalterin in [S4], obgleich die Hündin Rexi nicht das Potenzial hat, moralische Akteurin werden zu können. Springt sie jemanden an, dann wird die Halterin, die eine falsche Annahme über das Verhalten der Hündin hat und sie auf einer Straße bzw. auf Gehwegen unangeleint laufen lässt, zur moralischen (und vielleicht auch rechtlichen) Verantwortung gezogen, wenigstens dann, wenn etwas passiert, das nicht passieren darf. Ob der Killbot 3000 das Potenzial hat, moralischer Akteur werden zu können oder nicht, ist für die Haftbarmachung von jemandem, der mit seiner Zulassung, Entwicklung, Finanzierung und Nutzung zu tun hat, in seinem jetzigen Stadium unerheblich. Sicher ist: Für das Abfeuern genau des Schusses in [S6] sind die Zulassenden, Entwickelnden, Herstellenden, Finanzierenden und Nutzenden von Killbot 3000 nicht kausal verantwortlich. Sie sind aber hinsichtlich dieses Artefakts kausal dafür verantwortlich, dass es es überhaupt gibt und dass es so ausgestattet ist, dass es Schüsse abfeuern kann. Und sie sind kausal verantwortlich dafür, etwas in die Welt gebracht zu haben, das vielleicht Nichtvorhersehbares tut: Mit diesem Risiko mussten die Zulassenden, Entwickelnden, Herstellenden, Finanzierenden und Nutzenden kalkulieren. Das Risiko kannten sie. Sie mussten es kennen. Schlimmer noch: Sie kalkulierten fahrlässig angesichts von höchster Unsi-

cherheit. Daraus können sie sich nicht winden. Und wie im Fall des zweijährigen Kindes in [S2], das alleingelassen wird, sind die Genannten – in welchem Maße jeweils, ist eine schwierige Frage, weil viel mehr Hervorbringende als die Eltern in [S2] eine Rolle spielen – moralisch und, je nach Konstellation und Schwere dessen, was der Killbot 3000 tut, auch rechtlich dafür verantwortlich.

Schluss

In dem vorliegenden Beitrag haben wir dafür plädiert, die These der Verantwortungslücke im Zusammenhang mit opaken Systemen künstlicher Intelligenz zurückzuweisen: Ihr zufolge sind Aktionen dieser Systeme und deren Folgen solche, für die nichts und niemand moralisch (und eventuell auch rechtlich) verantwortlich gemacht werden kann. Wir haben zu zeigen versucht, dass dies nicht der Fall ist, und zwar aufgrund der Eigenart dieser Systeme, die weder Naturereignis sind, noch schicksalhaft auftauchen, noch moralische Akteure sind; und weil es Systeme sind, mit deren Entwicklung, Herstellung und Nutzung die Zulassenden, Entwickelnden, Herstellenden, Finanzierenden und Nutzenden wissen müssen, welches Risiko sie eingehen. Ihre Aktionen und deren Folgen wie Naturereignisse ansehen zu wollen oder wie die Handlungen von moralischen Akteuren und deren Resultate, beachtet die besondere Art des Artefakts nicht, das hier agiert.

Wer als Verantwortungsträgerin oder Verantwortungsträger selbstlernende, autonome Systeme künstlicher Intelligenz in die Welt bringen möchte, um bestimmte Zwecke zu erfüllen, tut gut daran, das Problem der Opazität dieser Systeme anzugehen, um sie beherrschbarer und vorhersehbarer, kurzum: kontrollierbarer zu machen, und somit auch die eigene moralische Verantwortung für ihr Sein und Wirkenkönnen nicht fälschlicherweise zu negieren, sondern anzuerkennen und produktiv mit ihr umzugehen. Solche Systeme transparenter zu machen, wie es im Kontext der sogenannten *erklärbaren künstlichen Intelligenz* versucht wird (Doran et al. 2017, Zednik 2021), ist hierfür gewiss dann der richtige Ansatz, wenn man überhaupt an dem Projekt festhalten möchte, solche Systeme zu entwickeln und in die Welt zu bringen.

Literaturverzeichnis

Ashrafian, H. (2015). Artificial Intelligence and Robot Responsibilities: Innovating Beyond Rights. *Science and Engineering Ethics* 21, 317–326. https://doi.org/10.1007/s11948-014-9541-0

Champagne, M. & Tonkens, R. (2015). Bridging the Responsibility Gap in Automated Warfare. *Philosophy & Technology* 28, 125–137. https://doi.org/10.1007/s13347-013-0138-3

Doran, D., Schulz, S., & Besold, T. R. (2017). *What Does Explainable AI Really Mean? A New Conceptualization of Perspectives.* https://doi.org/10.48550/ARXIV.1710.00794

Heidbrink, L., Langbehn, C. & Loh, J. (Hrsg.) (2017). *Handbuch Verantwortung.* Springer.

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, *6*(3), 175–183. https://doi.org/10.1007/s10676-004-3422-1

Misselhorn, C. (2018). Roboterethik. In N. Arnold & T. Wangermann (Hrsg.), *Digitalisierung und Künstliche Intelligenz. Orientierungspunkte.* Konrad Adenauer Stiftung, S. 96–109.

Santoni de Sio, F., & Mecacci, G. (2021). Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them. *Philosophy & Technology*, *34*(4), 1057–1084. https://doi.org/10.1007/s13347-021-00450-x

Tigard, D. W. (2021). There Is No Techno-Responsibility Gap. *Philosophy & Technology*, *34*(3), 589–607. https://doi.org/10.1007/s13347-020-00414-7

van de Poel, I., Royakkers, L. & Zwart, S. D. (Hrsg.) (2018). *Moral Responsibility and the Problem of Many Hands.* Routledge.

Zednik, C. (2021). Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence. *Philosophy & Technology*, *34*(2), 265–288. https://doi.org/10.1007/s13347-019-00382-7

Schöpferische Algorithmen?

Zum Zusammenhang von Weltverhältnis, Negativität und künstlerischem Schaffen

Christina Kast

„Ungeheuer ist viel. Doch nichts Ungeheurer, als der Mensch."
(Sophokles, Antigone)

Ein schöpferischer Algorithmus scheint per definitionem ein Widersinn. Der Algorithmus ist eine automatisierte Schrittfolge, eine Formel, zur Lösung eines Problems mittels der Durchführung von einfachen oder komplexen Berechnungen. Sein Tun bewegt sich im Bereich des klar Bestimmbaren: Er rechnet, kombiniert und schließt auf Grundlage von eingespeisten Datenmengen (vgl. Lenzen 2020, S. 29 f.) Das Schöpferische berührt dementgegen die Welt des Unbestimmten und Unberechenbaren. In der Kunst als Inbegriff von Schöpfung herrscht das Fantastische, Inspirative und Unwägbare. Beide Sphären stehen für sich, und mag der sogenannten intelligenten Maschine auch bedenkenlos eine höhere Recheneffizienz als dem Menschen zugebilligt werden, ist es fraglich, ob sie Zugang zur Welt des Schönen und Erhabenen hat.

Der Widersinn eines schöpferischen Algorithmus scheint sich in der Gegenwart indes in Wohlgefallen aufzulösen: Man glaubt sich der Verwirklichung des Traumes der kreativen Maschine so nah wie nie, denn Algorithmen malen, dichten und komponieren. Die Kunst als eine der „letzten Bastionen des Menschen" (Lenzen 2020, S. 65), die es gegen die Künstliche Intelligenz zu verteidigen gibt, ist – so scheint es – im Fall. Mehr noch: Man erhofft sich von der intelligenten Maschine nicht allein künstlerisches Potenzial, sondern spricht ihr ein das menschliche übersteigendes Vermögen zu, so sie schneller, effektiver und quantitativ mehr künstlerisches Output erzeugen kann – die menschliche Schaffenskraft bleibt scheinbar wenig staunenswert angesichts einer KI, die 5000 Choräle im Bach-Stil an nur einem Tag produziert. Und so lauschen Musikkenner algorithmischen Neu-Kompositionen im Stile Bachs oder staunen über die maschinelle Vollendung der „Unvollendeten" Beethovens oder Schuberts, während Kunstliebhaber von KI erzeugte Gemälde zu Höchstpreisen kaufen.

Vor diesem Hintergrund eröffnet sich das entscheidende philosophische Problemfeld da, wo zu entscheiden ist, ob Künstliche Intelligenz lediglich Werkzeug eines menschlichen Künstlers ist oder ob sie selbst als Künstler, d. h. als originär schaffender Akteur, angesehen werden kann. Wäre die KI in erstem Falle lediglich Hilfsmittel und Werkzeug des Künstlers – diesem gleichsam zur Hand – so müsste ihr im zweiten Falle das Vermögen zu künstlerischer Schaffenskraft zugesprochen werden. Das heißt: Die Maschine wäre darin eigenständiger, autonom schaffender Künstler, der nicht mehr von einer ihn lenkenden Instanz wie dem Menschen bedingt und abhängig ist.

Kann Künstliche Intelligenz Künstler sein? Der vorliegende Beitrag wendet sich gegen die Zuschreibung autonomer maschineller Künstlerschaft. Die Maschine, so ist zu zeigen, kann nicht Künstler sein. Der für die Untersuchung maßgebende Gedanke ist, dass eine so geartete Zuschreibung von schöpferischer Potenz den grundlegenden Umstand vernachlässigt, dass Kunst stets Ausdruck von Weltverhältnis und Welterfahrung des Künstlers ist. Der Künstler malt, komponiert und dichtet, um die von ihm wahrgenommene und erfahrene Welt im Kunstwerk sichtbar zu machen – darin ist Kunst genuin menschliche Lebensäußerung. Der Künstler ist zunächst Mensch und als lebendiges Wesen, als Lebewesen, bestimmt von seiner exzentrischen Stellung in der Welt sowie seiner Leiblichkeit, die je für sich Bedingung der Möglichkeit von Wahrnehmung, Weltverhältnis und Welterfahrung sind. Die Sphäre des Lebendigen ist algorithmisch nicht zu erreichen ebenso wenig wie das, was den Künstler im Speziellen auszeichnet, nämlich sein Vermögen zum Anders-Sehen der Dinge, welches seinen Ursprung in der Erfahrung von Negativität hat – in Entfremdung, Schmerz und leidvollem Zerwürfnis mit der Welt.

Der geschilderte Gedankengang wird in drei Schritten entfaltet: (1) Es wird zunächst ein Überblick über die Grundzüge der Debatte zur kreativen Künstlichen Intelligenz gegeben, um nach einer kurzen Skizzierung der Implikationen und Annahmen einer Befürwortung von kreativer KI zur Kritik überzugehen, welche sich im Rahmen des Beitrags auf den Umstand fokussiert, dass (2) der Künstler zunächst Mensch ist und Kunst demnach als menschliche Lebensäußerung zu verstehen ist. Dazu werden die Konzepte der exzentrischen Positionalität nach Helmuth Plessner sowie der Leiblichkeit nach Maurice Merleau-Ponty erläutert, um vor diesem Hintergrund der Zusammenhang von Weltverhältnis, leiblicher Wahrnehmung und Kunst zu erörtern. (3) Es folgt abschließend die Darlegung der Bedeutung der Negativität für den Prozess künstlerischer Schöpfung.

Die Vorstellung des autonomen KI-Künstlers und ihre Grenzen

Der Traum von der künstlichen Kreativität ist kein neuer. Er ist untrennbar von der Vorstellung, menschliches Denken sei maschinell simulierbar: Ist der Geist algorithmisch fass- und darstellbar, so muss auch das Phänomen menschlicher Kreativität aufzuschlüsseln und darzustellen sein. Versuche, eine algorithmische Theorie der Kreativität zu entwickeln und eine computergenerierte Kunst zu erzeugen, gehen zurück bis in die 1950er Jahre (vgl. dazu Rauterberg 2021, S. 14 ff.). Zwar finden sich bereits hier erste kybernetische Skulpturen und sogenannte stochastische Gedichte, doch kann von einem breitenwirksamen Interesse an der KI-Kunst erst seit dem Aufstieg der digitalen Technik gesprochen werden. Mehr Künstler denn je nutzen Künstliche Intelligenz als Werkzeug und Inspiration.

Für den angestrebten Aufstieg der KI vom „Spielgefährten des Künstlers“ hin zur „Künstlerin eigenen Rechts“ (Rauterberg 2021, S. 15 f.) entscheidend waren Fortschritte im Forschungsfeld des sogenannten *Deep Learning*, mittels dessen man die Möglichkeit einer eigenständig kreativen KI für gegeben glaubt. *Deep Learning*, ein neuerer Teilbereich des maschinellen Lernens, ist ein Meilenstein in der Fortentwicklung der Künstlichen Intelligenz. Intelligente Systeme zeichnen sich maßgeblich dadurch aus, dass sie fähig sind, selbstständig zu lernen. Mittels *Deep Learning* ist maschinelles Lernen ohne menschliches Eingreifen möglich, was bedeutet, dass die Maschine ihre Fähigkeiten eigenständig fortentwickeln kann, indem die neuronalen Netze neue Verknüpfungen formieren. Grundlage bildet das Ableiten und Extrahieren von Mustern aus eingespeisten Datenmengen, wie beispielsweise praktiziert bei der Bild- oder Spracherkennung (vgl. Lenzen 2020, S. 37 f.).

Ein Blick auf die gegenwärtigen Debatten zeigt, dass die Versuche, künstliche Kunst hervorzubringen, sowohl zahlreich als auch in allen Bereichen ästhetischen Schaffens vorzufinden sind, so in Musik, Malerei und Literatur: Der von Gaetan Hadjeres im Jahr 2017 entwickelte KI-Generator *DeepBach* verspricht so, über bloße Imitation und Reproduktion hinaus, neue Choräle im Stile Bachs selbst zu komponieren. Grundlage dazu ist ein maschineller Lernvorgang, der eine Generalisierung des Kompositionsprozesses möglich machen soll. Im Bereich der Malerei steht seit 2016 das KI-System *AICAN* für die selbstständige, d. h. autonome Generierung von Kunstwerken. Dem System wird von seinen Schöpfern eine Form von Kreativität zugesprochen, insofern es trainiert wurde, Neues hervorzubringen (vgl. Rauterberg 2021, S. 45). Dies gelingt, indem generierte Bilder vom Algorithmus ausgeschlossen werden, sobald sie bereits existierenden Bildern – *AICAN* wurde mit etwa 80.000 Bildern der Kunst gefüttert – zu sehr ähneln. 2018 gelang es einer Deep-Learning-KI eigene Sonette zu erzeugen, die denjenigen Shakespeares gleichen. Nur in

das Projekt involvierte Experten waren in der Lage, zwischen Original und KI-Lyrik zu unterscheiden.

Diese schlaglichtartige Beleuchtung einzelner Beispiele soll an dieser Stelle genügen. Für den Fortgang der Argumentation entscheidend sind die grundlegenden Implikationen, die aus einer Befürwortung der Möglichkeit von originärer KI-Kunst herauszulesen sind. Zweierlei Tendenzen zeichnen sich ab: Entweder steht das Kunstobjekt, die von der KI generierte Kunst, im Mittelpunkt der Diskussion, oder das kunstschaffende Subjekt, die Kunst generierende KI.

Im ersten Falle wird ein maschinell erzeugtes Kunstwerk als originäres anerkannt, wenn es für ein Produkt menschlicher Kreativität gehalten wird, d. h. nicht als KI-Kunst enttarnt wird. Ist dies der Fall, so hat sie im Sinne des Turing-Tests den Beweis erbracht, dass KI schöpferisch sein kann (Rauterberg 2021, S. 53). Zweierlei ist hierbei anzumerken: (1) Die Anerkennung eines algorithmisch erzeugten Kunstwerkes als Kunst erfolgt nicht durch die Maschine selbst, sondern kann nur durch den Menschen vollzogen werden. Die KI vermag dies nicht zu leisten, da sie – anders als der Mensch – nicht um das weiß, was sie tut. Nicht die KI bestimmt, was Kunst oder wer Künstler ist – also was sie selbst ist – sondern wiederum der Mensch. Allein dieser Umstand gäbe Anlass zu fragen, ob ein Künstler ein solcher sein kann, ohne von sich oder von dem, was er hervorbringt, zu wissen. (2) Diese Überlegung führt uns wiederum zu der Frage, nach welchem Maßstab hier ein Kunstwerk als ein solches bewertet wird. Augenscheinlich ist insbesondere die Wirkung, der unmittelbare sinnlich-ästhetische Genuss, maßgebend. Erscheint das Gehörte, Gesehene oder Gelesene als schön, so wird es in den Stand des Kunstwerks – sein Schöpfer in den des Künstlers – gehoben. Was Kunst ist, liegt in der Debatte um eine schöpferische KI also nicht mehr in der Sache selbst, einem Wesenskern, sondern unterliegt dem Urteil des Kunst-Konsumenten. Die Unabschließbarkeit der Diskussion über das Wesen der Kunst wird so zum Anlass genommen, den Kern zu verflüssigen und zu relativieren (so bei Rauterberg 2021, S. 17) – der Zeitgeist bestimmt, was Kunst und wer Künstler ist. Problematisch erscheint dies in zweierlei Form: Zum einen ließe sich dieses Argument beliebig ausweiten, insofern auch andere Disziplinen wie die Philosophie oder menschliche Tugenden wie die Gerechtigkeit nicht abschließend in eine Definition gebannt werden können. Zum anderen wird das Urteil über Kunst zur Frage der Nachfrage, die fortan den Kunstcharakter festlegt – der Massengeschmack lenkt, die Kunst wird zur Ware.

Subjektorientierte Argumentationen stellen die Frage nach dem Künstler bzw. dem künstlerischen Subjekt dezidiert in den Mittelpunkt der Diskussion (vgl. bspw. Volland 2018). Hier wird das Problem verhandelt, ob dem algorithmischen System Kreativität zugesprochen werden kann. Die Kreativität wird zum Maßstab der Anerkennung der KI als autonomem Künstler, zentral ist der Gedanke, Kunst müsse Neues und Unkonventionelles hervorbringen.

Künstler sei, wer Bestehendes überschreite. Auch dies lässt sich in groben Zügen problematisieren: Das Urteil ob der Neuheit des Kunstwerks kann wiederum nur der Mensch fällen, ebenso wie die Bestimmung dessen, wer und was kreativ ist. Der Algorithmus weiß nicht um das Neue in der Kunst ebenso wenig wie um das, was er bildlich, literarisch oder musikalisch erzeugt. Spricht man ihm Subjektstatus zu, so unter völliger Ausblendung des Umstands, dass Kunst sich primär aus dem Weltbezug des Künstlers nährt: Dies ist ihre Möglichkeitsbedingung. Diese ist nicht zu imitieren oder künstlich zu erzeugen, da sie aus einem genuin menschlichen Verhältnis zur Welt resultiert – und dieses ist exzentrisch und leiblich.

Der Künstler als Mensch: Exzentrizität und Leiblichkeit

Der Mensch in seiner exzentrischen Stellung zur Welt

In seinem Hauptwerk „Die Stufen des Organischen und der Mensch" entfaltet Helmuth Plessner eine philosophische Anthropologie, innerhalb derer der Mensch aus der Kontinuität mit allem Lebendigen heraus gedeutet wird. Das den Menschen gegen die Pflanzen- und Tierwelt Abgrenzende und Auszeichnende sieht Plessner in dessen „exzentrischer Positionalität". Mit dem Begriff der Positionalität umschreibt Plessner zunächst den Umstand, dass alles Lebendige im Gegensatz zu Nicht-Lebendigem in kontinuierlichem Austausch mit seiner Umwelt steht (Plessner 2016, S. 180). Positionalität als Form alles Lebendigen äußert sich zentrisch – beim Tier – oder aber exzentrisch – beim Menschen. Das Tier hat anders als die Pflanze ein Zentrum, seinen Organismus, und kann so als im Verhältnis zu seiner Umwelt unabhängig bezeichnet werden. Als zentral erweist sich dabei, dass das Tier zwar in der Lage ist, zwischen sich und seiner Umwelt zu differenzieren, aber mangels einer reflexiven Rückbezogenheit nicht um seine eigene Position in ihr weiß. Das Tier ist eins mit seinem Leib, mit seiner Mitte, und geht auf im „Hier-Jetzt" (Plessner 2016, S. 305) seiner Positionalität: Das Tier erlebt, doch es weiß nicht von seinem Erleben (Plessner 2016, S. 360). Dementgegen weiß der Mensch um sich, er ist sich seiner selbst bewusst, was bedeutet, dass er nicht in seiner eigenen Mitte aufgeht. Er ist seine Mitte, steht jedoch durch sein Selbstbewusstsein auch außerhalb von ihr; er ist in sich und zugleich nicht in sich: Er ist ex-zentrisch: „Er lebt und erlebt nicht nur, sondern er erlebt sein Erleben" (Plessner 2016, S. 364). Exzentrische Positionalität ist eine Lebensform, die ihrer eigenen Mitte ansichtig sein kann und nicht mehr in sich ruht.

Die Existenz des Menschen ist darin „auf Nichts gestellt" (Plessner 2016, S. 365): Das bedeutet zunächst, dass Mensch-Sein genuin Weltverhältnis ist, ein sich permanent ins Verhältnis zur Welt setzen müssen. Die „Hälftenhaf-

tigkeit“ (Plessner 2016, S. 385) des Menschen, seine natürlich instabile Lage, versetzt ihn in anhaltende existentielle Unruhe. Der Mensch existiert in anhaltender Spannung, denn „der Mensch lebt nur, indem er ein Leben führt“ (Plessner 2016, S. 384). Er muss sich erst zu dem machen, was er ist, und das stets aufs Neue: Der Mensch, so Plessners Konsequenz, ist von Natur aus ein Kulturwesen (Plessner 2016, S. 385). Die Kultur ist gleichsam die zweite Natur des Menschen. In ihr findet das menschliche Ringen um die Herstellung der eigenen Mitte, des Gleichgewichts inmitten seiner existentiellen Gleichgewichtslosigkeit, Ausdruck: „Als exzentrisches Wesen nicht im Gleichgewicht, ortlos, zeitlos im Nichts stehend, konstitutiv heimatlos, muß er ’etwas werden’ und sich das Gleichgewicht – schaffen“ (Plessner 2016, S. 385).

Mag die Exzentrizität auch Ausdruck von immerwährender Lebensnot sein, so ist sie gleichermaßen Quelle der Weltoffenheit und Geistigkeit des Menschen: Als exzentrisches Wesen können ihm Selbst und Welt gegenständlich werden und als ein zu Betrachtendes und zu Gestaltendes vor das eigene Bewusstsein treten. Seine Gebrochenheit ist *conditio sine qua non* für seine Abgehobenheit von der Bedingtheit des Tieres. Die ambivalente Brüchigkeit der menschlichen Lebensform ist für Plessner unaufhebbar: Eine finale Heilung des Bruchs, eine Vervollständigung des Menschen – wenn überhaupt möglich – käme seiner Abschaffung gleich. Die Installierung einer Mitte, die nicht mehr instabil und in Permanenz zu schaffen ist, bedeutet Verlust des Menschlichen in dem, was ihn vor allen Lebewesen auszeichnet: Geist und Kultur. Kultur ist in ihren Erscheinungsformen, so auch in der Kunst, Inbegriff des menschlichen Sich-Zur-Welt-Verhalten-Müssens. Insofern ist der Mensch und nur der Mensch Kulturwesen. Kultur als zweite Natur des Menschen ist aus ihm hervorgewachsen, aus Lebensnot angesichts der radikalen existentiellen Offenheit.

Folgt man Plessners Logik, so wird deutlich, wie fragwürdig die Annahme ist, die Maschine könne Schöpfer von Kunst oder Kultur sein, ist sie doch zunächst im Wesentlichen selbst Ergebnis und Ausdruck der menschlichen Exzentrizität – Geschöpf und nicht Schöpfer, Folge und nicht Ursprung. Dass die Kultur genuin an die menschliche Lebensform gebunden, ja Lebensäußerung des Lebewesens Mensch ist, macht zudem offenbar, worin der zentrale Mangel der Maschine liegt: Ihr fehlt das grundlegende Kriterium des Schöpferischen, nämlich das Vermögen, sich zur Welt und zu sich selbst ins Verhältnis setzen zu können – und zu müssen, erwächst dieses doch organisch aus dem von der Daseinsbewältigung getriebenen Menschen. Die Technik hat in diesem Sinne kein Verhältnis zur Welt, keine genuine Bindung an die Welt, die ihn zur Selbst- oder Weltauslegung bewegen und befähigen könnte.

In seiner Tiefe erlaubt Plessners Kulturbegriff keine Reduktion der Kultur in ihren Erscheinungen auf eine Sammlung von Daten, wie es beispielsweise Holger Volland unternimmt, wenn er darüber nachsinnt, dass Algorithmen durch maschinelles Lernen zu „lebendigen Subjekten“ werden und „kulturelle Selbstständigkeit“ erlangen könnten (Volland 2018, S. 23). Das Vermögen

dazu, so Volland, ergebe sich aus der Sammlung von riesigen Datenmengen, anhand derer Algorithmen lernen können, Muster herauszulesen und daraus zu begreifen, was Kultur sei (Volland 2018, S. 18), denn: Die Daten seien gleichsam die gesellschaftliche DNA (ebd., S. 23), welche die maschinelle Intelligenz entschlüsseln und kopieren könne. So kämen Algorithmen zur Gestaltung ihrer eigenen Kultur. Mehr noch: Mit Blick auf die Kunst werde eine „vollständige Datenbank der Weltkunst" immer greifbarer (ebd., S. 18) – ein allwissender Algorithmus wäre dann in Vollands Sinne, so ist zu schließen, ein vollkommener Künstler.

Die Gleichsetzung von Kultur – und Kunst – mit einem Gemenge von Daten, wie wir sie exemplarisch bei Volland finden, gründet in einer spezifischen Vorstellung vom Menschen, in einem szientistischen Menschenbild. Darunter fallen der Naturalismus und der Funktionalismus, welche den Menschen beiderseits als Summe seiner Daten begreifen, da er als reduzierbar auf neurale Prozesse bestimmt wird (vgl. Fuchs 2020, 10 f.). So führt der Naturalismus alle Phänomene auf naturwissenschaftliche Erklärungen zurück; Subjektivität, Geist und Bewusstsein gelten ihm als Produkt physiologischer Vorgänge, konkret neuronaler Prozesse. Der Funktionalismus seinerseits deutet Bewusstseinsphänomene als Prozesse neuronaler Informationsverarbeitung. Diese ist mittels künstlicher Systeme simulierbar, da der Geist gleich seiner Funktion ist, die in Datenverarbeitung und Output besteht. In der Folge eines solchen Menschenbildes wird der Geist als digitalisierbar und in der Folge als von Algorithmen imitierbar verstanden – sei es in Hinblick auf Intelligenz oder Kreativität. Nimmt man jedoch den Menschen, dem Szientismus entgegen, in seiner Ganzheit als Lebewesen, so wird dies zweifelhaft, da man in Anschlag zu bringen hat, dass der Mensch nicht nur Geist, sondern in seiner Gesamtheit primär Leib ist.

Leiblichkeit, Wahrnehmung und Welterfahrung

Als Lebewesen ist der Mensch vor allem eines: leibliches Wesen. Mit dem Konzept der Leiblichkeit ist die psychophysische Einheit des Menschen umschrieben, welche nebst der philosophischen Anthropologie insbesondere von der Phänomenologie gegen Positionen in Stellung gebracht wird, die den Menschen für sich sowie sein Verhältnis zur Welt in zwei Teile zerfallen lassen – in Geist und Körper, Bewusstsein und Wirklichkeit, Innen- und Außenwelt. Die phänomenologische Kritik richtet sich explizit gegen diese Dualismen der abendländischen Philosophie, die zumeist mit dem Primat des Geistes einhergehen und in objektivistisch-szientistischen Positionen wiederkehren (vgl. Zahavi 2018, S. 33 f).

Mit seiner Leibphänomenologie sucht Maurice Merleau-Ponty den Grundstein für eine Philosophie zu legen, die mit allen irrtümlichen Dualismen

sowie der Entwertung des Leibes im Verhältnis zum Bewusstsein aufräumt. In seiner *Phänomenologie der Wahrnehmung* wird der Leib als das dem Bewusstsein Vorgängige bestimmt, ohne diesen jedoch in Opposition zum Bewusstsein zu bringen. Im Vordergrund steht die Einheit beider, insofern Merleau-Ponty das Bewusstsein selbst als leiblich bestimmt: Es wohnt dem Leib ein und gehört ihm an. Als psychophysische Einheit ist auch der Leib seinerseits nicht isoliert von der ihn umgebenden Außenwelt: Er ist immer schon in Verbindung zur Welt zu denken, da ein lebendiger Organismus nicht an der Grenze des Körpers Halt macht, sondern in Wechselwirkung mit der Umwelt steht. Der Leib ist für den Menschen Medium zur Welt, erst durch ihn hat er Welt. Diesen grundlegenden Weltbezug bezeichnet Merleau-Ponty als leibliches Zur-Welt-Sein. (Merleau-Ponty 1966, S. 15 ff.)

Man kann daraus schließen, dass der Leib keine Nebensächlichkeit, kein beliebig austauschbarer Träger des Bewusstseins ist. Bewusstsein ist im Leib inkorporiert und als leibliches Phänomen immer schon durchdrungen von und versenkt in der Wirklichkeit, die der Leib zunächst wahrnehmend erfährt. Die primäre Welterfahrung ist die sinnlich-wahrnehmende Erfahrung der Wirklichkeit, aus der diese stets von Neuem erschlossen und konstituiert wird. Ein so beschaffenes leibliches Bewusstsein bildet keine von ihm unabhängig existierende, objektive Wirklichkeit ab. Ebenso wenig vollzieht es eine subjektive und willkürliche Konstruktion der Wirklichkeit. Das Verhältnis des menschlichen Subjekts zur Welt, von Bewusstsein zur Wirklichkeit, lässt sich nach Merleau-Ponty nicht erschöpfend in einem reinen Subjektivismus oder Objektivismus erfassen, da beide Theorieformen den Vermittlungscharakter der leiblichen Wahrnehmung zwischen Subjekt und Objekt, Mensch und Welt, vernachlässigen. Wahrnehmung, so heißt es bei Merleau-Ponty, sei gerade nicht wie das Licht eines Scheinwerfers, „der die Gegenstände in stets derselben Weise beleuchtet", sondern sei „Wirkursache des von ihm zur Erscheinung gebrachten" (Merleau-Ponty 1966, S. 47).

Was bedeutet dies nun für die Frage nach der Möglichkeit algorithmischer Schöpferkraft? Wahrnehmende Erfahrung der Welt steht am Anfang aller Künste – ob visuell oder akustisch, vermittelt der Künstler uns die Welt, wie sie sich ihm zeigt; wie er sie erfährt und wie sie sich ihm erschließt. Diese Voraussetzung kann nicht maschinell eingeholt werden, denn:

Das leibliche Zur-Welt-Sein des Menschen macht subjektive Wahrnehmung und Erfahrung von Welt überhaupt erst möglich. Die leibliche Existenz des Menschen ist nicht vernachlässigbar angesichts der leiblichen Vermitteltheit des menschlichen Weltbezugs, was deutlich macht, dass Wahrnehmung und Erfahrung von Welt nicht erschöpfend als Daten und Information, die von außen eingespeist und verarbeitet werden, zu deuten sind, wie szientistische Menschenbilder suggerieren: Die Deutung von Subjektivität, Geist und Bewusstsein als physikalisches Geschehen, dessen Funktionslogik auf algorithmische Vorgänge zurückgeführt werden könne – die grundlegende Voraussetzung

zung für den Anspruch einer Künstler-KI – ist zu kurz gegriffen. Hier stellt sich das Bewusstsein als Daten- und Informationsverarbeitung eines souverän funktionierenden Apparates – des Gehirns – dar, Leiblichkeit und Existenz werden zum grundsätzlich verzichtbaren Appendix. In der Konsequenz erscheint der Geist als unabhängig, d. h. loslösbar von Letzteren; eine Überzeugung, die beispielsweise in der Debatte um *Mind Uploading* manifest wird. Von dieser Warte aus erscheint die endgültige Loslösung des Geistes von seiner leiblich-existentiellen Bedingtheit als eine Verbesserung – nicht zuletzt da erst hier eine ungestörte Funktionsweise eines „reinen Geistes" möglich zu werden scheint (vgl. Fuchs 2020, S. 71–117). Eine solche Optimierung scheint auch in Hinblick auf den oder die autonome*n KI-Künstler*in erwartet zu werden: Er oder sie kennt keine kennt keine Schaffenskrise, keine Selbstzweifel, keine Krankheit – er funktioniert und produziert (vgl. Volland 2018, S.35).

Die leibliche Vermitteltheit von Welt impliziert darüber hinaus, dass die Wirklichkeit nicht erschöpfend in Daten erfassbar ist, dass Bewusstsein nicht einfach eine außerhalb von ihm bestehende, objektive Welt abbildet. Es selbst ist von dieser Welt durchdrungen und maßgeblich an der Konstituierung der Welt beteiligt. Wahrnehmung ist eine Praxis, die eine Anteilnahme an der Welt verlangt, eine Bewusstseinsleistung, die nicht allein eine vorgefundene Welt spiegelt, sondern in einem steten Prozess hervorbringt (vgl. Schürmann/Kast 2022). In diesem Sinne ist Wahrnehmung, so bspw. das Sehen, nicht als rein physiologischer Akt zu verstehen, sondern immer schon im Zusammenspiel mit weiteren Bewusstseinsleistungen wie bspw. dem Erinnern, dem Urteilen usw. zu denken. Es gibt kein reines Sehen, sondern immer nur ein Sehen von etwas als etwas. Sehen und Einsehen mögen analytisch trennbar sein, im aktiven lebendigen Vollzug sind sie jedoch als Einheit zu verstehen (vgl. Schürmann/Kast 2022). In Hinblick auf die Kunst bedeutet dies, dass wir in ihr und durch sie die Welt, wie sie sich in den Augen des Künstlers zeigt, zu sehen bekommen. Das Kunstwerk ist wahrnehmbare – für die bildenden Künste sichtbare – Manifestation der sich im Unsichtbaren vollziehenden Welterschließung des Künstlers.

Dem Algorithmus kann vor diesem Hintergrund weder Wahrnehmung noch Erfahrung von Welt, wie durch den Menschen durch- und erlebt, zugeschrieben werden. Ihm fehlt die Bedingung der Möglichkeit dazu, der Leib, und folglich die Verstricktheit in die und Durchdrungenheit von der Welt. Die Maschine ist „weltlos". Damit aber fehlt ihr die Grundvoraussetzung für die Schaffung eines Kunstwerks, für Schöpfertum – dieses ist nur möglich in den Lebensvollzügen des Menschen; nur der Mensch kann Künstler sein. Sein Verhältnis zur Welt, sein Zur-Welt-Sein, ist nicht künstlich herstellbar, da Ausdruck seiner Leiblichkeit.

Der Mensch als Künstler: Anders-Sehen und Negativität

Der Künstler ist Mensch – doch nicht jeder Mensch ist Künstler. Insofern ist zu klären, was das schöpferische Subjekt vor den anderen Menschen auszeichnet. Wenn Kunst als Materialisierung einer Sichtweise von Welt zu verstehen ist, wenn in der Kunst also die Welt, wie sie dem Künstler erscheint, offenbar wird, so muss im Blickwinkel des Künstlers etwas Auszeichnendes liegen, etwas, das ihn vom gewöhnlichen Blick auf die Dinge abgrenzt. Das heißt: Der Künstler sieht die Dinge anders und bringt sein Anders-Sehen der Welt mittels der ästhetischen Ausdrucksformen zum Vorschein. Die Kunst ist ihm „Medium der Unterbrechung intersubjektiv autorisierter Wahrnehmungsgewohnheiten" (Schürmann 2008, S. 211).

Das künstlerische Anders-Sehen der Welt gründet in einem spezifischen Weltverhältnis. Folgt man Helmuth Plessner, so ist die Fremdheit entscheidend dafür, dass die Welt als eine andere erscheint: „Wir nehmen nur das Unvertraute wahr. Um anschauen zu können, ist Distanz nötig" (Plessner 2020, S. 169). Fremdheit entsteht durch den Verlust der Vertrautheit, durch das Heraustreten aus dem Selbstverständlichen. Wer im Gegebenen aufgeht, vermag dieses nicht zu betrachten. Nur derjenige, dem das Gegebene als das Selbstverständliche abhandenkommt, vermag sich zu eben diesem ins Verhältnis zu setzen – es bewusst zum Gegenstand seiner Betrachtung zu machen. Die Fremdheit ist Bedingung, um überhaupt wahrzunehmen, zu sehen, gleichsam zur ersten Anschauung zu kommen. Die Dinge anders zu sehen, bedeutet folglich, sie gleichsam erstmals, neu sehen zu können.

Die „Kunst des entfremdeten Blicks" (Plessner 2020, S. 170) ist, so Plessner, eine schmerzliche Gabe, mag sie auch Voraussetzung aller echten Anschauung und allen Verstehens sein. Die den Blick verändernde Fremdheit ist mehr als eine epistemische Komponente, sie wurzelt in der Existenz des Menschen, im Menschen als Existenz. Und so bedeutet das Fremdwerden der ihn umgebenden Welt für den Künstler mehr als nur eine rein spielerische, abstrakte Neu-Perspektivierung der Welt. Es geht einher mit einer ganzheitlichen Erfahrung des leidvollen Bruchs mit der Welt, der den von jeder Norm abweichenden Zugang zur Welt möglich macht: „Der Schmerz ist das Auge des Geistes" (Plessner 2020, S. 172).

Aus diesen Überlegungen lässt sich schließen, dass der Künstler nicht zuletzt in Auseinandersetzung mit dem Unverfügbaren leben und schaffen muss: Die Welt erfährt er als etwas, das sich ihm in ihrer ursprünglichen Selbstverständlichkeit schmerzhaft entzieht, ihn in seinen Grundfesten erschüttert und verwundet. Diese Negativität ist ihm jedoch unentbehrlich, um zu schaffen – um die Welt in seinem Kunstwerk in neuem Lichte zu erfassen. Negativität sei hier im Sinne Byung-Chul Hans verstanden als Inbegriff des Anderen, des sich Entziehenden, d. h. des Umstands, dass Welt ein Gegenüber ist, das nicht in die

eigene Logik gezwungen und verfügbar gemacht werden kann (vgl. Han 2016); als etwas, das widerfährt und das man erleidet; als etwas, das sich dem eigenen Zugriff entwindet und zunächst in Unverständnis zurücklässt.

Das heißt: Der menschliche Künstler produziert nicht einfach Neues, welches er aus einer gleichsam interesselosen Ansammlung von Erfahrungen und Sichtweisen zusammensetzt. Er gebiert es, bringt es gleichsam zur Welt, insofern es aus ihm erwächst, mit allem dazu gehörigen Leiden und Schmerz, die seinem spannungsvollen Verhältnis zur Welt inhärent sind. Anders zu sehen bedeutet, anders zu sein – anders zu leben. Dementgegen kennt die Maschine nur die reine Positivität, insofern sie über Daten ausschließlich verfügt und diese beherrscht – sie kennt keine Ungewissheit, keinen Zweifel, keine Verwirrung. Sie hat nur Antworten, keine Fragen, da sie die „Negativität des Bruchs“ (Han 2021, S. 50) nicht an sich erfahren und erleben kann. Sie vermag es in der Konsequenz ebenso wenig Künstler wie Philosoph zu sein. Das Ende der Negativität wäre das Ende der Kunst, ob unter Menschen oder Maschinen.

Es ist nichtsdestotrotz gerade die Abwesenheit von Negativität, die als Garant für die maschinelle Überlegenheit über den Menschen angeführt wird (vgl. Volland 2018, S. 35). Bar jeden menschlichen Unbills sammelt die Maschine Daten, erkennt Muster und „kreiert“ bzw. produziert „störungsfrei“ sogenannte Kunst. Für sich vollends unbedarft, führt der Algorithmus seine Aufgaben aus. Dass hier ein Fortschritt bzw. eine Optimierung des Kunst-Schaffens vermutet wird, offenbart, dass Existenz – ebenso wie Lebendigkeit und Leiblichkeit – nicht als Ingredienz von Schöpfertum gewertet wird, sondern bestenfalls als vernachlässigbar, im Grunde aber als störend. Es ließe sich hier freilich fragen, worin konkret die Vorzüge einer „schmerzfrei“ und „weltlos“ produzierten maschinellen Kunst bestehen sollen. Für den Grundgedanken des Beitrags festzuhalten ist jedoch, dass auch hier das, was konstitutiv für den menschlichen Geist und seine Schöpfungen ist, als Laune der Natur, die vom Menschen behoben werden kann und soll, übergangen wird.

Fazit

Können Algorithmen schöpferisch, die Maschine Künstler sein? Intention des Beitrags war es, diesen Anspruch kritisch zu hinterfragen. Dabei wurde gezeigt, dass die Möglichkeit autonomer Künstlerschaft von intelligenten maschinellen Systemen zu verneinen ist, da Schöpfertum etwas voraussetzt, das genuin menschlich und künstlich nicht imitiert werden kann: im Allgemeinen ein Weltverhältnis sowie Wahrnehmung von Welt, die in der Stellung des Menschen als Lebewesen gründen, konkret in seiner Exzentrizität und Leiblichkeit; im Speziellen das künstlerische Vermögen zum Anders-Sehen der Welt, wel-

ches aus der Erfahrung von Negativität, wie Entfremdung, Schmerz und Leid entspringt.

Dass der Zeitgeist dessen ungeachtet der Möglichkeit schöpferischer Algorithmen euphorisch entgegenfiebert, ist, wie gezeigt wurde, auf ein grundlegend szientistisches Bild des Menschen zurückzuführen, welches Lebendigkeit, Leiblichkeit sowie Existenz zum vernachlässigbaren Appendix eines datenbasierten Geistes degradiert. Doch wird die Nivellierung von Mensch und Maschine zudem durch einen gedankenlosen und leichtsinnigen Sprachgebrauch befördert und gefördert. Wenn beteuert wird, KI sammle Erfahrungen, analysiere Strukturen und löse sich von der Vergangenheit, um so – gleich dem Menschen – auf dieser Basis Neues und Überraschendes zu schaffen (vgl. Rauterberg 2021, S. 26), so wird sprachlich etwas suggeriert, das nicht existent ist: nämlich ein lebendiges und seiner selbst bewusstes Subjekt, das in seiner von Brüchen durchsetzten Existenz, von der Welt durchwirkt, eben diese in der Kunst zur Anschauung bringt.

Literatur

Fuchs, T. (2020). *Verteidigung des Menschen. Grundfragen einer verkörperten Anthropologie.* Suhrkamp.

Han, B.-C. (2016). *Die Austreibung des Anderen. Gesellschaft, Wahrnehmung und Kommunikation heute.* Fischer-Verlag.

Han, B.-C. (2021). *Undinge. Umbrüche der Lebenswelt.* Ullstein Verlag.

Lenzen, M. (2020). *Künstliche Intelligenz. Fakten, Chancen, Risiken.* Beck Verlag.

Merleau-Ponty, M. (1966). *Phänomenologie der Wahrnehmung.* De Gruyter.

Plessner, H. (2016). *Die Stufen des Organischen und der Mensch.* Suhrkamp.

Plessner, H. (2020). Mit anderen Augen. In Ders. (Hrsg.), *Mit anderen Augen. Aspekte einer philosophischen Anthropologie.* Reclam, S. 164–182.

Rautenberg, H. (2021). *Die Kunst der Zukunft. Über den Traum der kreativen Maschine.* Suhrkamp.

Schürmann, E. (2008). *Sehen als Praxis. Ethisch-ästhetische Studien zum Verhältnis von Sicht und Einsicht.* Suhrkamp.

Schürmann, E. & Kast, C. (2022). Sehen. In J. Sigmundt (Hrsg), *Handbuch Kunstphilosophie.* UTB (im Erscheinen).

Volland, H. (2018). *Die kreative Macht der Maschinen. Warum künstliche Intelligenzen bestimmen, was wir morgen fühlen und denken.* Beltz Verlag.

Zahavi, D. (2018). *Phänomenologie für Einsteiger.* UTB.

Wehrhaft oder wahrhaft? – Politische Ansätze und normative Paradoxien der Regulierung von Internetinhalten in liberalen Demokratien

Wolf J. Schünemann

1. Einleitung

Dieser Beitrag richtet seinen Blick auf die politischen Ansätze der Regulierung von Internetinhalten und allgemeiner der Internetkontrolle in liberalen Demokratien. Dabei soll nicht allein der Wandel von Regulierungspraktiken im Hinblick auf die Online-Kommunikation betrachtet werden, sondern insbesondere auch die Entwicklung der Regulierungsdiskurse in Bezug auf digitale Medien anhand der darin erkennbaren legitimatorischen Strategien. Folgt man aktuell vorherrschenden Begründungsfiguren für die Regulierung von Online-Kommunikation in liberalen Demokratien, insbesondere im Hinblick auf die so genannte Desinformation, dann soll die gesteigerte Regulierungsaktivität in erster Linie dem Schutz gesellschaftlicher Autonomie und demokratischer (Informations-)Freiheit dienen. Doch sind die derzeit diskutierten oder bereits implementierten Gegenmittel tatsächlich geeignet, um die liberale Demokratie und ihre grundlegenden Werte gegen ihre inneren und äußeren Feinde zu verteidigen, oder stellen sie womöglich selbst eine Gefahr für unsere freiheitliche Ordnung dar? Mit dieser zentralen Frage des Beitrags wird ein grundlegenderer Konflikt zwischen Meinungsfreiheit und informationeller Autonomie auf der einen Seite und dem Schutz der Demokratie auf der anderen Seite berührt, der gerade in der Tradition der deutschen staatsrechtlichen Betrachtung prominent diskutiert worden ist und mit dem Begriff der wehrhaften oder auch der streitbaren Demokratie beschrieben wird (vgl. Thiel 2003). Dieser Konflikt scheint heute im Spannungsverhältnis zwischen demokratisch begründeten Einschränkungen von Informationsflüssen im Netz durch Staat und Diensteanbieter und der Informationsautonomie des Individuums in aktualisierter Form

auf und bringt – so möchte ich argumentieren – erhebliche normative Paradoxien hervor. Einige von ihnen werde ich in den folgenden Abschnitten herausarbeiten. Dazu werde ich zunächst auf den aktuell sichtbaren Wandel demokratischer Medienpolitik hinsichtlich Internet und Digitalisierung hinweisen. Daran anschließend möchte ich einen Überblick über verschiedene interdisziplinäre Forschungsstränge geben, die sich mit der Internetkontrolle aus unterschiedlichen Blickwinkeln befasst haben. Im Hauptteil beschreibe ich den Wandel demokratischer Rechtfertigungsmuster für die Online-Inhalteregulierung an den Beispielen Frankreichs und Australiens. Im Ergebnis versuche ich die folgende These zu begründen:

Mit der gesteigerten Bedrohungswahrnehmung im Hinblick auf Cyberangriffe auf die Wahlinfrastruktur und ausländische Desinformationskampagnen ist der Schutz demokratischer Souveränität zu einer vorrangigen Legitimationsstrategie für regulatorische Eingriffe in die Online-Kommunikation avanciert. Dies trägt auch in liberalen Demokratien zu einer Zunahme von Regulationsaktivität auf verschiedenen Ebenen (Staat und Intermediäre) mit größerer gesellschaftlicher Akzeptanz bei. Als normativ paradoxe Folgen sind Beeinträchtigungen der Informationsfreiheit sowie der grenzüberschreitenden Online-Kommunikation zu erwarten.

2. Die Regulierung von Internetinhalten und das Ende des Exzeptionalismus

Entgegen den libertären Visionen, die mit der Internetentwicklung in ihren frühen Jahren verbunden waren (Barlow 1996, Johnson & Post 1996), stellt die Regulierung von Internetinhalten kein absolutes Novum dar, sondern hat sich auch in liberalen Demokratien schon früh als rechtsstaatliche Notwendigkeit erwiesen und musste teils gegen Widerstände der an manche Freiheitserzählungen der Pionier*innen anknüpfenden, frühen Intermediäre durchgesetzt werden (am Bsp. Yahoo in Goldsmith & Wu 2006, allgemein: Timofeeva 2006). Allerdings ließen die Regulierungsakteure in liberalen Demokratien dabei eine systemkonforme medienpolitische Zurückhaltung erkennen, die geradezu als eine Funktionsbedingung für das liberale Demokratiemodell gelten kann. Die Meinungs- und Informationsfreiheit gehört zu den Grundwerten der Demokratie und muss gerade in Form freier Medien, d. h. Presse, Rundfunk, Fernsehen und auch Internetmedien verwirklicht werden. Eingriffe in diese Freiheit und damit auch die Informationsautonomie des Individuums, das auf die mediale Informationsversorgung angewiesen ist, sind normativ prekär. Oder – wie es der Medienwissenschaftler Douglas Vick (2001) ausdrückt –: „When democratic countries regulate media content, they engage in a

precarious balancing act". Demokratische Regierungen bewegen sich bei der Medienregulierung also auf einem schmalen Grat. Die normativen Voraussetzungen ihres Herrschaftsmodells ergeben eine Voreingenommenheit gegen strenge Regulierungen oder einen inhärenten „bias against control" (McQuail 2008). Aufgrund weltpolitischer Konstellationen und der US-amerikanischen Dominanz in der Internetentwicklung, aufgrund regulatorischer Trends in der zweiten Hälfte des 20. Jahrhunderts und natürlich auch aufgrund der technischen Charakteristika des Internets ist die normativ begründete medienpolitische Zurückhaltung für das neue Medium und seine Träger bislang besonders deutlich ausgeprägt. Dies schlägt sich insbesondere in den verschiedentlich kodifizierten Haftungsprivilegien für Internetanbieter nieder, die in der Regel eben nicht für die Inhalte haften, die Nutzer*innen auf den bereitgestellten Plattformen einstellen. Sowohl der US-amerikanische Communications Decency Act (Section 230) aus dem Jahr 1996 als auch die europäische E-Commerce-Richtlinie aus dem Jahr 2000 sehen ein entsprechendes Privileg vor. Gegenüber anderen Medienträgern, die strengeren Regeln im Hinblick auf ihre redaktionelle Verantwortung unterliegen, resultiert daraus eine Ungleichbehandlung, die auch als *Internetexzeptionalismus* bezeichnet wird (Fried 2021).

In den vergangenen Jahren mehren sich die Anzeichen, dass dieser Exzeptionalismus bröckelt. Regulierungsakteure in liberalen Demokratien stellen ihre Zurückhaltung gegenüber regulierenden Eingriffen zunehmend infrage und setzen geltendes Recht und/oder normative Wertvorstellungen durch strengere Regeln und effektive Sanktionen durch. Darauf weisen etwa die jährlich erscheinenden Berichte von Freedom House, die Freedom on the Net Reports, hin. Im aktuellen Bericht wird im zehnten Jahr in Folge ein Rückgang der Internetfreiheit global festgestellt. 26 der insg. 65 untersuchten Staaten verzeichneten Abzüge für die entsprechenden Indexwerte, darunter auch liberale Demokratien, die für die Abwärtstrends mitverantwortlich sind. Gerade die Phänomene Hasskriminalität und Hetze sowie Falschnachrichten oder sog. Desinformation sind die aktuell vorherrschenden und augenscheinlich auch in der demokratischen Welt stärker akzeptierten Rechtfertigungsgründe für Eingriffe in die Informations- und Kommunikationsfreiheit. Dies wird insbesondere auch im pandemischen Kontext offenbar. Die Wahrnehmung einer regelrechten „Infodemie" – so der von der WHO selbst geprägte Ausdruck – und entsprechende Warnungen dienen als Legitimationsstrategien, um die Regulierung von Internetinhalten zusätzlich zu verschärfen (Freedom House 2020).

Schaut man auf die Gesetzgebung, so kann das von Deutschland vor wenigen Jahren beschlossene Netzwerkdurchsetzungsgesetz (NetzDG) als Beispiel für den unterstellten Trend dienen. Mit seiner Verabschiedung im Vorfeld der letzten Bundestagswahl 2017 ist die Bundesrepublik international als Vorreiterin in Sachen demokratischer Internetregulierung in Erscheinung getreten. Das Gesetz verpflichtet Anbieter großer sozialer Netzwerke (über 2 Mio. Nutzer*innen) rechtswidrige Inhalte binnen 24 Stunden zu entfernen. Strittige

Inhalte müssen binnen sieben Tagen überprüft werden. Zudem gelten weitere Regeln zu Beschwerdemanagement und Transparenz. Das Gesetz wurde erst im Frühjahr 2021 novelliert. So ist nun etwa auch die direkte Meldung der Urheber*innen strafrechtlich relevanter Inhalte beim BKA vorgeschrieben. Zwar wurde das NetzDG vielfach als Schwellenübertritt zur Zensur kritisiert, es hat aber international auch viel Beachtung gefunden. Andere Länder haben sich an wesentlichen Vorgaben des Gesetzes orientiert. So ist etwa das 2019 verabschiedete sog. ‚loi Avia' zur Bekämpfung von Hasskriminalität in Frankreich klar vom NetzDG inspiriert. Auch Elemente des im Dezember von der EU vorgelegten Entwurfs eines Digital Services Act tragen Züge des NetzDG (Schünemann 2021a). Schließlich geraten die Regelungen der Section 230 des Communications Decency Act, einschließlich des Haftungsprivilegs, selbst in den USA zunehmend unter Druck. Im zurückliegenden US-Präsidentschaftswahlkampf wurde die zentrale Regelung von beiden konkurrierenden Seiten infrage gestellt. Unter Präsident Trump war noch im Oktober 2020 ein Reformvorschlag eingebracht worden. Eine Reform unter der Präsidentschaft Bidens und damit die Veränderung einer Grundregel des US-amerikanisch geprägten Internetökosystems in der laufenden Legislaturperiode kann als wahrscheinlich gelten. Zahlreiche Gesetzesvorschläge sind bereits von beiden parlamentarischen Lagern eingebracht und in Kongressausschüssen beraten worden.

3. Zwischen medienpolitischer Zurückhaltung und Souveränität – ein kursorischer Überblick über den Forschungsstand

In diesem Beitrag geht es um das Spannungsverhältnis zwischen Informationsautonomie und dem Schutz demokratischer Freiheiten angesichts bedrohlicher Phänomene ‚im Netz'. Mithin rücken auch Fragen staatlicher Regulierungsfunktion und Souveränität in den Blick. Der in diesem Abschnitt zu leistende Überblick über den Forschungsstand soll eng auf diese Fragestellungen hin ausgerichtet werden und kann sowohl aus Platzgründen als auch aufgrund seiner interdisziplinären Behandlung nicht erschöpfend sein. Der kursorische Überblick kann ausgehend von der erwähnten medienpolitischen Zurückhaltung liberaler Demokratien bei den medienwissenschaftlichen Ursprüngen dieser These ansetzen. Hier ist der normativ begründete „bias against control" (McQuail 2008) als Grundregel etabliert und auf die Internetentwicklung angewendet worden (Schejter & Han 2011). Zugleich haben führende Vertreter*innen jedoch stets darauf hingewiesen, dass Massenmedien immer vielfältig politisch institutionalisiert seien (Jarren 2007). Ein nachholender,

zweistufiger Institutionalisierungsprozess, wobei regulatorische Innovation auf den Medienwandel erst reagieren muss, stellt aus medienhistorischer Sicht mithin die Regel und nicht die Ausnahme dar (Stöber 2007, S. 107). So reiht sich das, was aktuell als Ende des Internetexzeptionalismus diskutiert wird, womöglich ein in eine lange Reihe regulatorischer Anpassungsleistungen an Phänomene medialen Wandels. Aktuell scheinen die innovativen Selbst- und Ko-Regulierungsarrangements für eine Content-Regulierung und -Moderation mit und durch Plattformen (platform governance) und andere Anbieter sowie gerade das Zusammenspiel von staatlicher Regulation und Intermediären diese Anpassungsleistungen zu reflektieren und erhalten folglich zunehmend Beachtung im Feld (Flew et al. 2019, Katzenbach & Ulbricht 2019).

Auch die vergleichende Politikwissenschaft hat die medienpolitische Zurückhaltung von Demokratien aufgegriffen. Auf dieser Wahrnehmung gründet – gleichsam ex negativo – der lange Zeit und bis in die Gegenwart hinein nahezu ausschließliche Fokus des entsprechenden Forschungsstrangs (Hellmeier 2016, Keremoğlu & Weidmann 2020) auf die Kontrolltechniken und -praktiken autokratischer Regime: „It has become a truism to link censorship in cyberspace to the practices of authoritarian regimes“ (Deibert & Rohozinski 2010, S. 15). Demgegenüber gibt es nur eine kleine Zahl regimetypenübergreifender Vergleichsstudien (Stier 2017, Timofeeva 2006). In jüngeren Jahren, nicht zuletzt angesichts der in empirischen Vergleichsstudien wie etwa von Freedom House dokumentierten Trends, sind allerdings vermehrt auch liberale Demokratien in den Blick genommen worden. Busch et al. haben aus ihrer wiederum exklusiven Beschäftigung mit den neuen Regulierungsaktivitäten demokratischer Regime gar eine Art umgekehrter Diffusionshypothese abgeleitet: „Learning from autocracies“ (Busch 2017, Busch et al. 2018, Wright & Breindl 2013).

Demgegenüber reicht die rechtswissenschaftliche Literatur zum Thema besonders weit zurück und hat sich seit den Entstehungsjahren des Internets als Massenmedium mit dem Leitkonzept der Souveränität befasst. Dabei hat Rechtswissenschaftler*innen die Frage beschäftigt, welche Handlungsspielräume und Regulierungsoptionen staatliche Regulationsakteure im Umgang mit digitalen Medien überhaupt noch haben. Während die oben bereits zitierte libertäre Szene von Netzaktivist*innen nationalstaatlichen Regierungen sehr deutlich Legitimation und Kompetenz für die Internetregulierung absprach (Barlow 1996) und darin auch von namhaften Wissenschaftler*innen unterstützt wurde (Johnson & Post 1996), wurde der These von einem souveränen Cyberspace als einer neuen Form der Souveränität jenseits von Staatlichkeit ebenso deutlich widersprochen. So zeigten etwa Goldsmith und Wu (2006) in ihrem Buch „Who controls the Internet?“, wie staatliche Gesetzgeber und Gerichte durchaus in der Lage waren, rechtsstaatliche Normen und souveräne Gestaltungsansprüche trotz Widerständen von Plattformen und Intermediären durchzusetzen.

Souveränitätsvorstellungen unterschiedlicher Art haben sich gerade in den jüngeren Jahren vielfach verbreitet und sind in politischen Diskussionen zur Regulierung von Internet und Digitalisierung ausgesprochen virulent. Forderungen nach technologischer Souveränität, Datensouveränität, informationeller Souveränität finden sich prominent in Regierungsdokumenten verschiedener Nationalstaaten, aber auch Strategiepapieren der EU und der USA. Eine Reihe rezenter Studien hat diese Trends theoretisch und diskursanalytisch aufgearbeitet (Pohle & Thiel 2019, Thiel 2019) oder an aktuellen Beispielen illustriert (Möllers 2021, Schünemann 2021b). Als einer der führenden Internet-Governance-Experten hat Milton Mueller diesen Trend für die internationale Diskussion um die politische Regulierung des Internets bereits in seiner Schrift „Will the Internet fragment?" (2017) pointiert zum Ausdruck gebracht:

„The call for protection was often couched in the language of sovereignty. There are now explicit demands for data sovereignty, technological sovereignty, or various other labels for some kind of governmental or jurisdictional overlay on information networks. They come from Brazil, Germany, the UK, and the European Commission as well as from China, Russia, Saudi Arabia, and Iran. The message is the same: realign states and cyberspace" (ebd., S. 12).

Forderungen nach größerer Souveränität in digitalpolitischen Fragen und insbesondere im Hinblick auf liberale Demokratien als Regimetyp lassen sich wiederum aus einem anderen Forschungsstrang ableiten. In den Internationalen Beziehungen und strategischen Studien ist in den vergangenen Jahren insbesondere das Thema von Desinformationskampagnen aufgegriffen worden. Daraus ergeben sich breit diskutierte Bedrohungsszenarien im Rahmen asymmetrischer Konfliktkonstellationen. Liberale Demokratien steckten demnach geradezu in einem Desinformationsdilemma. In dem Maße, in dem Desinformationskampagnen aus dem nicht-demokratischen Ausland zu einer asymmetrischen Bedrohung für liberale Demokratien im politischen Systemwettbewerb würden, gereichten die liberalen Grundsätze der Demokratie den demokratischen Staaten zum strategischen Nachteil in internationalen Konfliktsituationen. Vor diesem Hintergrund könnte sich die lange Zeit über verteidigte und hochgehaltene, normativ begründete Zurückhaltung liberaler Demokratien gegenüber regulatorischen Eingriffen in die Medien- und insbesondere die Internetfreiheit in eine wesentliche Schwäche und größere Verwundbarkeit verwandeln (vgl. Goldsmith & Russell 2018).

Die Regimetypendifferenz aufgreifend, wird zudem darauf hingewiesen, dass Autokratien, die hinter den Kampagnen zu vermuten seien, besser in der Lage seien, ihre eigene „digitale Souveränität" zu schützen (Omand 2018, Pope 2018). Gerade die schädlichen Effekte, die Desinformationskampagnen nach verbreiteter Auffassung auf zurückliegende Wahlkämpfe in demokratischen Staaten gehabt haben (Brexit-Referendum, Wahl Donald Trumps zum US-Präsidenten etc.), haben die asymmetrische Verwundbarkeit von Demokratien

zum Thema auch jenseits strategischer Studien werden lassen (Jamieson 2018). Gerade aus dem angenommenen auswärtigen Ursprung hat sich gegenüber interner strategischer Kommunikation die sicherheitspolitische Deutung einschließlich der Diskussion von Gegenmaßnahmen bis hin zu militärischen Operationen gespeist. Um neue Legitimationsstrategien demokratischer Regierungen im Kontext der Desinformationsbekämpfung soll es in den folgenden Abschnitten gehen.

4. Die Legitimation von Inhalteregulierung – eine revidierte Typologie von Rechtfertigungsmustern

Wie in den vorangegangenen Abschnitten bereits deutlich geworden ist, reicht das Spannungsverhältnis von Kommunikations- und Informationsfreiheit auf der einen Seite und dem Schutz der Demokratie vor ihren Feinden auf der anderen Seite weiter und ist grundlegender, als es die aktuell vorherrschende Beschäftigung mit Phänomenen der Digitalisierung nahelegen mag. Tatsächlich geht das „Problem" der Demokratie mit der freien Rede weiter zurück als die Digitalisierung. Entsprechend sind die grundlegenden Problemdimensionen schon durch frühere wissenschaftliche Werke in der Politik-, Rechts- und Medienwissenschaft erhellt worden. Unter diesen Arbeiten ragt Cass Sunsteins Buch von 1995 „Democracy and the problem of free speech" (Sunstein 1995) hervor, weil es die normativ herausfordernde Begründungsnotwendigkeit, aber auch die möglichen und legitimen Rechtfertigungsweisen einer demokratischen Inhalteregulierung vor dem Hintergrund des US-amerikanisch geprägten, in kritischer Bezeichnung so genannten „free speech absolutism" reflektiert und ausführlich darlegt. Für die folgenden Überlegungen stütze ich mich im Wesentlichen auf die für das Buch zentrale dreigliedrige Typologie von Rechtfertigungsweisen für Inhalteregulierung (ebd., S. 3–12 u. 173; s. auch Sunstein 2017). Als ersten Typ für die Regulierung freier Rede definiert Sunstein – zunächst kontraintuitiv – die *inhalte-neutrale* Rechtfertigung von Regulierungsentscheidungen („content-neutral"). Als allgemeines Beispiel dient ihm das Einkaufszentrum, in dem es prinzipiell untersagt sein kann, öffentliche Reden zu halten oder Pamphlete zu verteilen, ganz gleich, welchen Inhalts diese auch sein mögen. Ein Beispiel aus dem Bereich digitaler Informationsräume wäre das Verunstalten von Websites (das sog. defacement), etwa auch als Angriffsform im Bereich der Cybersicherheit oder subversiver Akt in Protestkontexten. Ein zweiter Typus fasst demgegenüber *inhalte-abhängige* Regulierung zusammen, die aber unabhängig von Ansichten und Meinungen ist, die durch die betreffenden Inhalte zum Ausdruck gebracht werden („content-based, but viewpoint neutral"). In diese Klasse fällt der Großteil der weithin als legitim

anerkannten Eingriffe in die Informations- und Kommunikationsfreiheit, etwa zur Bekämpfung gewisser Formen der Pornographie (vor allem Kinderpornographie), des Glücksspiels oder anderer krimineller Aktivitäten. Als legitime Regulationsziele stehen hier etwa der Jugendschutz, der Schutz der öffentlichen Gesundheit oder sicherheitspolitische Ziele im Vordergrund. Die Bekämpfung der benannten Inhalte erstreckt sich auch, und in zunehmendem Maße, auf digitale Informationsräume und Verbreitungswege, aber beschränkt sich nicht darauf. Ähnliches gilt für die ebenfalls digitalpolitisch virulenten Problemfelder der Hasskriminalität und des Extremismus. Allerdings machen kontextabhängig zu bewertende Grenzfälle im Hinblick auf diese Beispiele auch schon die teils fließenden Übergänge zu einem dritten Typ ersichtlich. Diesen bezeichnet Sunstein als meinungsbasiert („viewpoint-based"). Tatsächlich sind auch in liberalen, insbesondere in wehrhaften Demokratien (s. oben), bestimmte Inhalte aufgrund normativer Setzungen untersagt, weil gewisse Ansichten aus dem legitimen Meinungsspektrum ausgeschlossen und ihre Äußerung eingeschränkt oder untersagt ist. Ein prominentes Beispiel, das auch in der digitalpolitischen Regulierung bereits zu Kontroversen geführt hat, betrifft die Leugnung des Holocaust oder die Verherrlichung des NS-Regimes.

Gerade die Regulierung von Desinformation, wie wir sie aktuell verstärkt beobachten, fällt im Hinblick auf die Rechtfertigungsweisen in der Regel unter den dritten Typ. Denn dass für eine Großzahl von Kommunikaten zweifelsfrei und rechtssicher rein inhaltebasiert festgestellt werden könnte, dass es sich um eine zumindest in Teilen falsche Information handelt, muss als ausgesprochen fraglich erscheinen, von normativem Problem der Nominierung und der Machtfülle einer entsprechend beauftragten Instanz der Faktenprüfung ganz zu schweigen. Neben der Definition der Typen nimmt Sunstein selbst eine Reihung nach dem Grad der normativen Spannung zwischen Schutzzielen und informationeller Freiheit vor. Die Bewertung fällt wenig überraschend so aus, dass die inhalte-neutrale Rechtfertigung von Regulierung aus Sicht des Autors am wenigsten problematisch ist, die inhalte-basierte Rechtfertigung schon deutlich heikler, aber in vielen Fällen demokratisch legitimierbar sein kann, die meinungsabhängige Rechtfertigung von Regulierung in liberalen Demokratien indes besonders prekär und am schwierigsten zu legitimieren ist (Sunstein 2017, S. 208).

Wirft man einen genaueren Blick auf die Legitimationsstrategien demokratischer Inhalteregulierung im Hinblick auf Desinformation werden allerdings weitere Muster erkennbar, die sich nicht mehr gut mit dem von Sunstein geprägten Schema fassen lassen. Tatsächlich weisen aktuelle Bedrohungsszenarien im Hinblick auf die Verbreitung von Desinformation zumeist auf die Kampagnenförmigkeit sowie den ausländischen Ursprung der Bedrohung hin. Diese Annahmen gehen ein in die Heuristiken zur Erkennung schädlicher Kommunikate sowie die Begründung ihrer Regulierung. Vor diesem Hintergrund ist der von Sunstein entwickelten Typologie heute ein vierter Typ hin-

zuzufügen. Diesen möchte ich als *ursprungsbasiert* bezeichnen. Regulierungsnotwendigkeit und -praxis gegenüber Inhalten werden durch Hinweis auf den auswärtigen Ursprung von Kommunikaten gerechtfertigt, wobei im Regelfall von einer spezifischen Gefährdungsbeurteilung je nach vermutetem Ursprungsland auszugehen ist. Ursprungsbasierte Legitimation von Inhalteregulierung erwarten wir sowohl für staatliche Regulierung, Ko-regulierung sowie Selbstregulierungsarrangements, aber auch konkrete Moderationspraktiken, etwa auch durch die Intermediäre selbst. Im folgenden Abschnitt werden mit Frankreich und Australien zwei in verschiedener Hinsicht herausragende Beispiele demokratischer Online-Kontrolle gewählt, um aktuelle Entwicklungen der Legitimationsstrategien von Content-Regulierung zu illustrieren.

5 Exemplarische Legitimationsstrategien demokratischer Inhalteregulierung

5.1 Frankreich und die Regulierung von Desinformation im Wahlkampf

Frankreich ist ein liberaldemokratischer Staat mit einem hohen Grad an Medienfreiheit. Gleichwohl steht das Land für eine Tradition staatlicher medienpolitischer Interventionen (Humphreys 1996). Wie vergleichende Studien zur demokratischen Medienpolitik gezeigt haben, weist das Land im internationalen Vergleich eine hohe Aktivität im Bereich der Inhalteregulierung („censorship rate“) auf, wobei die entsprechende Statistik historisch klar hinter die Internetentwicklung zurückreicht (Newton & Artingstall 1994). Wenngleich auch Frankreich in den 1980er und 90er Jahren durch eine Phase der Liberalisierung des Mediensystems gegangen ist (Humphreys 1996), ist die französische Tradition des starken Staates doch auch in der Internetregulierung sichtbar geworden, etwa in frühen Auseinandersetzungen mit Content Service Providern um schädliche Inhalte, Urheberrecht oder Datenschutz (vgl. Goldsmith & Wu 2006, Solum 2009).

Mit dem *Loi infox*, dem Gesetz zur Bekämpfung von Falschnachrichten im Wahlkampf, einem im europäischen Kontext neuartigen Gesetz aus dem Jahr 2018, werden große soziale Netzwerke erstmals verpflichtet, die Verbreitung von Falschnachrichten in Wahlkampfzeiten (drei Monate bis Wahltermin) zu bekämpfen, einschließlich richterlicher Löschanordnungen. Nachdem er selbst im Präsidentschaftswahlkampf 2017 Opfer von Cyber-Attacken und damit verbundenen Desinformationskampagnen geworden war, initiierte die Regierung unter Präsident Macron das Gesetzgebungsverfahren. In der Begründung der Initiative brachte Macron insbesondere die Bedrohung demokratischer

Souveränität durch ausländische Akteure zum Ausdruck: „die liberale Demokratie selbst ist das Ziel“ (Macron 2018, Übers. d. Verf.). An einer anderen Stelle wird das Feindbild ersichtlich, das die Bedrohung ausmacht:

„Diese wachsende Verbreitung von Falschnachrichten geht heute vollkommen Hand in Hand mit der Faszination für ein illiberales Politikmodell, die ich auch deshalb erwähne, weil sie häufig von den gleichen Kräften unterstützt wird. Sie wird häufig von *denselben Mächten* benutzt, die sich über die angeblichen Schwächen der Demokratie und ihre große Liberalität lustig machen, über ihre Unfähigkeit zur Ordnung, zur Hierarchie, letztlich zur Anerkennung von Autorität“ (ebd.).

Mit der „illiberalen Vision“ und ihren Proponenten ergeben sich ein vages Feindschema: „Feinde der Demokratie“ (ebd.) und eine Bedrohung der Demokratie von außen als potenzielle ursprungsbasierte Rechtfertigung für Regulierung.

5.2 ‚FARA‘-Gesetzgebung in Australien

Auch Australien ist eine Demokratie mit einem hohen Grad politischer Freiheiten und Medienfreiheit. Allerdings gilt Australien gerade im Hinblick auf die Internetregulierung auch als Beispiel für einen relativ strengen Ansatz staatlicher Regulierung unter den liberalen Demokratien (Freedom House 2019). Bereits das nationale Telekommunikationsgesetz aus dem Jahr 1997 erlaubte es der Regierung, illegale Online-Dienste zu blockieren und den Zugang auf entsprechende Seiten einzuschränken. Über die lange geltende Rechtslage hinaus wurde in Reaktion auf das Christchurch-Attentat von 2019 das Strafgesetzbuch angepasst. Internetdienstanbieter wurden verpflichtet „abscheuliche Gewaltdarstellungen“ („abhorrent violent material“) zu löschen.

Auch im Hinblick auf die ursprungsbasierte Regulierung von Online-Kommunikation hat die australische Regierung im Dezember 2017 einen Reformvorschlag im Rahmen der nationalen Sicherheitsgesetzgebung unterbreitet. Der Vorstoß erfolgte in Reaktion auf Berichte über verdeckte Einflussoperationen, inklusive der Verbreitung von Desinformation. Die Reform orientiert sich am Muster des US-amerikanischen *Foreign Agents Registration Act* (kurz: FARA). Das ursprüngliche US-amerikanische FARA stammt aus dem Jahr 1938 und fällt in den Bereich der Geheimdienstregulierung. FARA-ähnliche Regeln haben in den vergangenen Jahren gerade im Hinblick auf die Regulierung von Online-Kommunikation eine erhebliche Konjunktur erfahren, nicht nur in autokratischen Staaten oder defekten Demokratien – über entsprechende Reformen in Russland (2012) und Ungarn (2017) wurde, insbesondere in Europa, sehr kontrovers diskutiert –, sondern zunehmend auch in liberalen Demokratien wie dem Vereinigten Königreich oder eben Australien.

FARA-ähnliche Gesetze sehen im Wesentlichen vor, dass ausländische Akteure, die mediale Inhalte verbreiten wollen, sich bei einer staatlichen Stelle registrieren müssen. Ihre Aktivitäten können so besser beobachtet werden, sie können bei Regelverstößen leichter sanktioniert werden. Werden sie zudem dazu verpflichtet, ihre Inhalte auf den ‚auswärtigen Ursprung' hin zu kennzeichnen, droht zudem eine Stigmatisierung und Diskriminierung im Rahmen öffentlicher Kommunikationsprozesse.

In der Rechtfertigungskommunikation der australischen Regierung werden sehr ähnliche Deutungsmuster erkennbar wie zuvor im französischen Beispiel. So machte etwa der damalige Premierminister Australiens, Malcolm Turnbull, das eigentliche Regulationsziel, nämlich die Bekämpfung von Desinformationskampagnen, durchgeführt von autokratischen Staaten in feindlicher Absicht, sehr deutlich:

„Authoritarian states have been literally manufacturing public opinion in order to hijack political discourse and tilt the decision-making landscape to their advantage.
[…] these methodologies have been turbocharged by cyber.
The very technology that was designed to bring us together, the internet, is being used as an instrument of division" (Turnbull 2017).

Die fundamentale Bedrohungswahrnehmung richtet sich also auf Desinformationskampagnen von außen. Wenngleich die Regierung eine vorgebliche Gleichbehandlung aller Akteure auswärtigen Ursprungs unterstreicht, wird in den Äußerungen des Premierministers und anderer Regierungsmitglieder doch klar ersichtlich, dass es um gewisse autokratische Staaten, aus australischer Perspektive an erster Stelle um China, geht. In diesem Zusammenhang wird auch die Annahme asymmetrischer Verwundbarkeit liberaler Demokratien gegenüber autokratischen Gegnern explizit hervorgehoben:

„We will not allow foreign states to use our freedoms to erode freedom; our open democracy to subvert democracy; our laws to undermine the rule of law" (Turnbull 2017).

Schließlich zeigen jüngste Berichte über die Praxis der Inhalteregulierung in der zurückliegenden nationalen Wahl 2019, dass über die staatliche Regulierung hinaus auch Plattformen wie Facebook ihre eigenen Maßnahmen der Content-Regulierung auf ausländische Desinformationskampagnen ausgerichtet haben, was angesichts der regierungsseitig gestützten Bedrohungswahrnehmung nicht verwundert, aber für eine zusätzliche faktische Diskriminierung ausländischer Inhalte im Sinne einer ursprungsbasierten Inhalteregulierung spricht.

6. Zusammenfassung und normative Paradoxien

In den vorangegangenen Abschnitten habe ich den eingangs besprochenen Paradigmenwechsel in der Regulierung von Online-Kommunikation durch liberale Demokratien, der sich in jeweils übertriebener Form – sowohl im Hinblick auf den Ausgangszustand als auch das realistische Entwicklungsszenario – als ‚Ende des Laissez-faire' oder als ‚Ende des Internet-Exzeptionalismus' bezeichnen ließe, anhand aktueller regulatorischer Debatten und Praktiken in verschiedenen Beispielländern illustriert. Ich habe dabei zum einen einen Schwerpunkt auf die Legitimationsstrategien und ihre Entwicklung gelegt, weil diese Hinweise sowohl auf grundlegendere Muster politischer Legitimation als auch die Wahrscheinlichkeit demokratischer Akzeptanz geben (vgl. hierzu Schünemann 2022). Zum anderen habe ich für Legitimationsstrategien und -praktiken auf rezente Muster und Ansätze fokussiert, die einem ursprungsbasierten Typ der Regulierung entsprechen. Gerade die in aktuellen Debatten virulente sicherheitspolitische Ausdeutung von demokratischen Krisenerscheinungen und insbesondere der Schutz der Demokratie vor ihren äußeren Feinden tragen als augenscheinlich erfolgreiche Legitimationsstrategien zum konstatierten Politikwandel bei.

Dieser Wandel hat – unabhängig von der politischen Bewertung – eine Reihe normativer Paradoxien zur Folge, die die Informationsautonomie von Nutzer*innen potenziell erheblich schwächen können, obwohl die Reformen mit dem Schutz der Demokratie gerade auf die institutionelle Absicherung politischer Autonomie zielen.

An erster Stelle steht ganz grundlegend die zu erwartende weitere Steigerung regulatorischer Maßnahmen und Einschränkungen der Informations- und Kommunikationsfreiheit. Dies geht einher mit einer Ermutigung von Intermediären, teilweise auch als konkrete Forderungen, tiefere und intransparentere Eingriffe in die Internetfreiheit vorzunehmen. Die beobachtbare Zunahme ursprungsbasierter Regulierung wird sich mit großer Wahrscheinlichkeit zulasten grenzüberschreitender Informationsflüsse auswirken. Schließlich werden souveränitätsbezogene Legitimationsstrategien autokratischer Regime, wie sie seit langem vorgetragen werden, durch gesteigerte Regulierrungsaktivitäten demokratischer Regierungen validiert. Dies kann in internationalen Konfliktkonstellationen um die Governance der Digitalisierung zu einer weiteren Schwächung liberaler Ordnungsvorstellungen führen und auch dadurch schädlich auf die individuelle Informationsautonomie als Bestandteil universeller Rechte in einer digitalisierten Umgebung wirken.

Literaturverzeichnis

Barlow, J.-P. (1996). A Declaration of the Independence of Cyberspace. https://projects.eff.org/~barlow/Declaration-Final.html [Zugriff: 13.03.2015].

Busch, A. (2017). Netzzensur in liberalen Demokratien. In A. Croissant, S. Kneip, & A. Petring (Hrsg.), *Demokratie, Diktatur, Gerechtigkeit. Festschrift für Wolfgang Merkel.* Springer VS, S. 331–352.

Busch, A., Theiner, P. & Breindl, Y. (2018). Internet Censorship in Liberal Democracies: Learning from Autocracies? In J. Schwanholz, T. Graham, & P.-T. Stoll (Hrsg.), *Managing Democracy in the Digital Age: Internet Regulation, Social Media Use, and Online Civic Engagement.* Springer International Publishing, S. 11–28.

Deibert, R. & Rohozinski, R. (2010). Control and Subversion in Russian Cyberspace. In R. Deibert, J. G. Palfrey, R. Rohozinski, & J. Zittrain (Hrsg.), *Access controlled. The shaping of power, rights, and rule in cyberspace. Information revolution and global politics.* MIT Press, S. 15–34.

Flew, T., Martin, F. & Suzor, N. (2019). Internet regulation as media policy: Rethinking the question of digital communication platform governance. *Journal of Digital Media & Policy,* 10(1), 33–50.

Freedom House (2019). Freedom on the Net 2019. The Crisis of Social Media. New York/Washington.

Freedom House (2020). Freedom on the Net 2020. The Pandemic's Digital Shadow. New York/Washington.

Fried, N. (2021). The Myth of Internet Exceptionalism: Bringing Section 230 into the Real World. *American Affairs,* V(2), 179–191.

Goldsmith, J. & Russell, S. (2018). Strengths Become Vulnerabilities. How a digital world disadvantages the United States in its international relations. In: Aegis Series, 1806, S. 1–22.

Goldsmith, J. & Wu, T. (2006). *Who controls the Internet? Illusions of a borderless world.* Oxford University Press.

Hellmeier, S. (2016). The Dictator's Digital Toolkit: Explaining Variation in Internet Filtering in Authoritarian Regimes. *Politics & Policy* 44(6), 1158–1191.

Humphreys, P. (1996). *Mass media and media policy in Western Europe. European Policy Research Unit series.* Manchester Univ. Press.

Jamieson, K. H. (2018). *Cyberwar. How Russian hackers and trolls helped elect a president; what we don't, can't, and do know.* Oxford University Press.

Jarren, O. (2007). Die Regulierung der öffentlichen Kommunikation. *Zeitschrift für Literaturwissenschaft und Linguistik,* 37(2), 131–153.

Johnson, D. R. & Post, D. (1996). Law and Borders: The Rise of Law in Cyberspace. *Stanford Law Review,* 48(5), 1367.

Katzenbach, C. & Ulbricht, L. (2019). Algorithmic governance. *Internet Policy Review,* 8(4). https://doi.org/10.14763/2019.4.1424

Keremoğlu, E. & Weidmann, N. B. (2020). How Dictators Control the Internet: A Review Essay. *Comparative Political Studies* 12(5). https://doi.org/10.1177/0010414020912278

Macron, E. (2018). Déclaration de M. Emmanuel Macron, Président de la République, sur la liberté de la presse, la lutte contre les fausses nouvelles, l'économie des médias et sur la relation entre les gouvernants et les médias, à Paris le 3 janvier 2018. https://www.vie-publique.fr/discours/204657-declaration-de-m-emmanuel-macron-president-de-la-republique-sur-la-li [Zugriff: 26.3.2020].

McQuail, D. (2008). *McQuail's mass communication theory. 5th ed., reissued.* SAGE Publications.
Möllers, N. (2021). Making Digital Territory: Cybersecurity, Techno-nationalism, and the Moral Boundaries of the State. *Science, Technology, & Human Values,* 46(1), 112–138.
Mueller, M. (2017). *Will the Internet Fragment? Surveillance, Cybersecurity and Internet Governance.* Polity Press.
Newton, K. & Artingstall, N. (1994). Government and Private Censorship in nine Western Democracies in the 1970 s and the 1980 s. In I. Budge & D. McKay (Hrsg.), *Developing democracy. Comparative research in honour of J. F. P. Blondel*. Sage, S. 297–321.
Omand, D. (2018). The threats from modern digital subversion and sedition. *Journal of Cyber Policy,* 3(1), 5–23.
Pohle, J. & Thiel, T. (2019). Digitale Vernetzung und Souveränität: Genealogie eines Spannungsverhältnisses. In I. Borucki, & W. J. Schünemann (Hrsg.), *Internet und Staat. Perspektiven auf eine komplizierte Beziehung.* Nomos, S. 57–80.
Pope, A. E. (2018). Cyber-securing our elections. *Journal of Cyber Policy,* 3(1), 24–38.
Schejter, A. M.& Han, S. (2011). Regulating the media: four perspectives. In D. Lévi-Faur (Hrsg.), *Handbook on the politics of regulation.* Elgar, S. 243–279.
Schünemann, W. J. (2022). Aufwärtskompatibel? Zur Bedeutung struktureller und doktrinärer Nationalismen für die digitale Konstellation, Zeitschrift für Politikwissenschaft,33(2), 293-311. DOI: 10.1007/s41358-021-00285-0.
Schünemann, W. J. (2021a). New horizontal rules for online platforms across Europe. A comment on the EU Commission's proposal for a Digital Services Act for DTCT partners and upstanders. https://dtct.eu/wp-content/uploads/2021/02/DTCT-TR1-DSA.pdf [Zugriff: 03. 03. 2021].
Schünemann, W. J. (2021b). Souveraineté numérique pour une société de l'information? L'Allemagne et le projet de normalisation de la politique numérique européenne, *Allemagne d'aujourd'hui, n° 236/avril-juin 2021.*
Solum, L. B. (2009). Models of Internet Governance. In J. Bing & L. A. Bygrave (Hrsg.), *Internet governance. Infrastructure and institutions*. Oxford University Press, S. 48–91.
Stier, S. (2017). *Internet und Regimetyp. Netzpolitik und politische Online-Kommunikation in Autokratien und Demokratien. Vergleichende Politikwissenschaft.* Springer Fachmedien Wiesbaden.
Stöber, R. (2007): Kommunikationsfreiheit und ihre Feinde. *Zeitschrift für Literaturwissenschaft und Linguistik* 37(2), 104–119.
Sunstein, C. R. (1995). *Democracy and the problem of free speech.* With a new afterword. 1. Free Press paperback ed. Free Press.
Sunstein, C. R. (2017). *#Republic. Divided Democracy in the Age of Social Media.* Princeton University Press.
Thiel, M. (Hrsg.) (2003). *Wehrhafte Demokratie. Beiträge über die Regelungen zum Schutze der freiheitlichen demokratischen Grundordnung.* Mohr Siebeck.
Thiel, T. (2019). Souveränität: Dynamisierung und Kontestation in der digitalen Konstellation. In J. Hofmann, N. Kersting, C. Ritzi & W. J. Schünemann (Hrsg.), *Politik in der digitalen Gesellschaft. Zentrale Problemfelder und Forschungsperspektiven.* transcript; Walter de Gruyter GmbH, S. 47–60.

Timofeeva, Y. (2006). *Censorship in cyberspace. New regulatory strategies in the digital age on the example of freedom of expression. 1. Aufl. Schriften zur Governance-Forschung, Band 6.* Nomos.

Turnbull, M. (2017). Speech introducing the National Security Legislation Amendment (Espionage and Foreign Interference) Bill 2017. https://www.malcolmturnbull.com.au/media/speech-introducing-the-national-security-legislation-amendment-espionage-an [Zugriff: 26. 3. 2021].

Vick, D. W. (2001). Exporting the First Amendment to Cyberspace: The Internet and State Sovereignty. In N. Morris, S. R. Waisbord & K. Nordenstreng (Hrsg.), *Media and globalization. Why the state matters.* Rowman & Littlefield, S. 3–19.

Wright, J., & Breindl, Y. (2013). Internet filtering trends in liberal democracies: French and German regulatory debates. *Internet Policy Review,* 2(2), https://doi.org/10.14763/2013.2.122

Algorithmen und Autonomie

Wir leben in einer Welt der algorithmischen Sortierung und Entscheidungsfindung. Mathematische Modelle kuratieren unsere sozialen Beziehungen, beeinflussen unsere Wahlen und entscheiden sogar darüber, ob wir ins Gefängnis gehen sollten oder nicht. Aber wie viel wissen wir wirklich über Code, algorithmische Strukturen und deren Wirkweisen? Der Band wendet sich den Fragen der Autonomie im digitalen Zeitalter aus einer interdisziplinären Perspektive zu, indem er Beiträge aus Philosophie, Erziehungs- und Kulturwissenschaft mit der Informatik verbindet.

Die Herausgeber*innen:
Prof. Dr. Dan Verständig, Professor für Erziehungswissenschaft mit dem Schwerpunkt Allgemeine Medienpädagogik, Universität Bielefeld
Dr. Christina Kast, wiss. Mitarbeiterin, Fakultät für Humanwissenschaften, Bereich Philosophie, Otto-von-Guericke-Universität Magdeburg
Janne Stricker, M.A., wiss. Mitarbeiterin, Fachbereich Medienpädagogik, Universität Bielefeld
Prof. Dr. Andreas Nürnberger, Professor für Data and Knowledge Engineering, Fakultät für Informatik, Institut für Technische und Betriebliche Informationssysteme, Otto-von-Guericke-Universität Magdeburg

ISBN 978-3-8474-2520-5



www.budrich.de

Titelbildnachweis: shutterstock.com